PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/435

A1

(11) International Publication Number: **WO 00/04142**

(43) International Publication Date: 27 January 2000 (27.01.00)

(21) International Application Number: PCT/AU99/00579

(22) International Filing Date: 16 July 1999 (16.07.99)

(30) Priority Data:
PP 4708 16 July 1998 (16.07.98) AU

(71) Applicants *(for all designated States except US)*: THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH [AU/AU]; The Bancroft Centre, 300 Herston Road, Herston, QLD 4029 (AU). THE CORPORATION OF THE TRUSTEES OF THE ORDER OF THE SISTERS OF MERCY IN QUEENSLAND [AU/AU]; Aubigny Place, Mater Misericordiae Hospitals, South Brisbane, QLD 4101 (AU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WILLIAMS, Stephanie, Jane [AU/AU]; 48 Stevenson Street, The Grange, QLD 4051 (AU). ANTALIS, Toni, Marie [AU/AU]; 527 Coronation Drive, Toowong, QLD 4066 (AU). McGUCKIN, Michael, Andrew [AU/AU]; 27 Leigh Street, Coorparoo, QLD 4151 (AU). GOTLEY, David, Charles [AU/AU]; 100 Birdwood Terrace, Auchenflower, QLD 4066 (AU).

(74) Agent: FISHER ADAMS KELLY; Level 13, AMP Place, 10 Eagle Street, Brisbane, QLD 4000 (AU).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.

(54) Title: MUCINS

(57) Abstract

prflated *MUC* nucleic acids are provided which correspond to a *Mucin* gene located on human chromosome 7q22, or on a mammalian chromosome structurally or functionally equivalent thereto, which *Mucin* gene is normally predominantly expressed in the colon. Also provided are diagnostic and therapeutic uses of isolated *MUC* nucleic acids, MUC polypeptides encoded thereby and anti–MUC mAb.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		

TITLE

"MUCINS"

FIELD OF THE INVENTION

THIS INVENTION relates generally to nucleic acids corresponding to mammalian Mucin genes, and to polypeptides encoded thereby. More particularly, the present invention provides isolated nucleic acids which correspond to Mucin regulatory genes that are predominantly expressed in the colon. These Mucin genes are associated with disease conditions including colorectal cancer, breast cancer, cystic fibrosis, respiratory diseases, inflammatory bowel disease, ulcerative colitis and Crohn's disease and/or any other conditions associated with aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins. In particular, the present invention provides methods for the diagnosis and therapy of the abovementioned disease conditions.

BACKGROUND OF THE INVENTION

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. This is particularly the case in cancer research. However, despite the effectiveness of this powerful technology, progress has been slow in developing effective recombinant DNA-derived therapeutic or diagnostic agents for cancers. One difficulty has been a lack of understanding of many cancers and other disease conditions. Regulatory genes are an important component of these complex regulatory mechanisms.

Cancer suppressor genes, for example, are regulators of cell growth and differentiation (Weinberg *et al.*, 1995, Ann. NY Acad. Sci. **758** 331). The paradigm for their role in cancer is that they are trans-acting and recessive at the cellular level; loss of one homologue has no effect on cell function and homozygous inactivation is required for carcinogenesis (Cavenee *et al.*, 1983, Nature **305** 779).

Colorectal cancers contribute to a major proportion of the

mortality and morbidity associated with cancer development. There is a particular need, therefore, to understand the complex regulatory mechanisms associated with colorectal cancers as well as cancers in anatomically adjacent regions.

The epithelial mucins are a family of secreted and cell surface glycoproteins expressed by epithelial tissues. They are characterised by a central polymorphic tandem repeat structure, which comprises most of the protein backbone, and a large number of O-linked carbohydrate side chains (Gum *et al.*, 1995, Biochem. Soc. Trans. **23** 795). The complex structure and large size of these molecules makes it difficult to characterise them using classical biochemical techniques. The genes are also difficult to clone because of their large size and the presence of GC-rich tandem repeats. Ten mucin genes have been identified; *MUC3*, *MUC4*, *MUC5AC*, *MUC5B*, *MUC6* and *MUC8* have been partially cloned and full-length cDNA clones are available for *MUC1*, *MUC2*, *MUC7* and *MUC9*.

Mucins are known to contribute to pathology in a number of epithelial diseases including cystic fibrosis (CF), inflammatory bowel disease (IBD) and adenocarcinomas. Gastrointestinal mucins which have been described to date include: the transmembrane mucins MUC1 and MUC4; the gel-forming mucins MUC2, MUC5AC and MUC6; and MUC3 which has an unclear structure and function.

As used herein, Mucin genes or isolated nucleic acids corresponding thereto will be expressed in italicized form as *MUC*. Mucin polypeptides will be expressed as MUC.

Immunohistochemical staining and Western blotting analysis with mature MUC1-specific antibodies revealed that MUC1 became ectopically expressed in colorectal tumours and levels were significantly higher in primary tumours of patients with metastases. Experimentally increased expression of gel-forming mucins resulted in increased metastasis in colon cancer cells in xenograft metastasis models (Ho *et al.*, 1995, Int. J. Oncol. **7** 913). Northern blot analysis has been employed to investigate

expression of *MUC1*, *MUC2*, *MUC3* and *MUC4* in paired normal and colonic tumour tissues and in nine colorectal cancer (CRC) cell lines (Ogata *et al.*, 1992, Cancer Res. **52** 5971). *MUC1* and *MUC4* were present in colonic mucosa with similar expression levels in carcinomas, but occasionally elevated levels of *MUC4* were apparent. Levels of *MUC2* and *MUC3* were decreased by varying degrees in the tumours of most patients. There was no apparent correlation between the expression of any mucin gene and the site, stage or histological type of tumour. All four mucin genes were expressed at low levels or not at all in the nine CRC cell lines under investigation; *MUC1* transcripts were detected in COLO205, *MUC2* and *MUC4* probes hybridised weakly to all nine cell lines, and *MUC3* expression was observed in five of the lines. Using a combination of *in situ* hybridisation and immunohistochemistry, Chang *et al* (Chang *et al.*, 1994, Gastroenterology **107** 28) also found *MUC2* and *MUC3* were downregulated in CRC. A more recent *in situ* hybridisation study found expression of *MUC2* and *MUC3* mRNA was markedly reduced in poorly, moderately and well-differentiated colorectal tumours but preserved in mucinous carcinomas (Weiss *et al.*, 1996, J. Histochem. Cytochem. **44** 1161). It is noted that *MUC3* is located on human chromosome 7q22, or an equivalent location on other mammalian chromosomes, and is primarily expressed under normal conditions in the small intestine (Shekels *et al.*, 1998, Biochem J. **330** 1301).

OBJECT OF THE INVENTION

The present inventors have realized that the Mucins constitute an incomplete family of genes and gene products implicated in a variety of disease conditions. Surprisingly, the present inventors have identified novel Mucin genes located on human chromosome 7q22, and isolated novel nucleic acids corresponding thereto. Furthermore, the present inventors have found that these novel Mucin genes are predominantly expressed in the colon, and may be involved in cancer of the large bowel, cystic fibrosis, breast cancer, inflammatory bowel disease, ulcerative colitis respiratory diseases and Crohn's disease and/or any other conditions associated with

aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins.

It is therefore an object of the invention to provide novel Mucin genes and isolated nucleic acids corresponding thereto.

SUMMARY OF THE INVENTION

The present invention is broadly directed to an isolated *MUC* nucleic acid which corresponds to a *MUC* gene located on mammalian chromosome 7q22, or on a mammalian chromosome structurally or functionally equivalent thereto, which *MUC* gene is normally predominantly expressed in the colon.

In a first aspect, the *MUC* gene of the present invention is *MUC11*. Accordingly, "a *MUC11* nucleic acid" means an isolated nucleic acid of the invention which corresponds to the *MUC11* gene.

Preferably, the isolated *MUC11* nucleic acid comprises a nucleotide sequence encoding an amino acid sequence which comprises SGLSEESTTSHSSPGSTHTTLSPASTTT (SEQ ID NO: 1).

More preferably, the isolated *MUC11* nucleic acid comprises a nucleotide sequence encoding the amino acid sequence according to SEQ ID NO:3.

Even more preferably, the isolated *MUC11* nucleic acid comprises a nucleotide sequence according to SEQ ID NO: 2.

In a second aspect, the *MUC* gene of the present invention is *MUC12*. Accordingly, "a *MUC12* nucleic acid" means an isolated nucleic acid of the invention which corresponds to the *MUC12* gene.

Preferably, the isolated *MUC12* nucleic acid comprises a nucleotide sequence encoding an amino acid sequence which comprises SGLSQESTTFHSSPGSTETTLAPASTTT (SEQ ID NO: 4).

More preferably, the isolated *MUC12* nucleic acid comprises a nucleotide sequence encoding the amino acid sequence according to SEQ ID NO:6.

Even more preferably, the isolated *MUC12* nucleic acid

comprises a nucleotide sequence according to SEQ ID NO: 5.

In a third aspect, the present invention resides in an isolated MUC polypeptide.

In one embodiment, the isolated MUC polypeptide has an amino acid sequence according to SEQ ID NO: 3, hereinafter referred to as a "MUC11 polypeptide".

In another embodiment, the isolated MUC polypeptide has an amino acid sequence according to SEQ ID NO:6, hereinafter referred to as a "MUC12 polypeptide".

In a fourth aspect, the present invention resides in an antibody specific for a MUC polypeptide (hereinafter referred to as an anti-MUC antibody).

Preferably, the anti-MUC antibody is selected from the group consisting of:-

(i) an anti-MUC11 IgM monoclonal antibody hereinafter referred to as M11.9; and

(ii) an anti-MUC12 IgM monoclonal antibody hereinafter referred to as M12.15.

In a fifth aspect, the present invention resides in methods of detecting a *MUC* gene, a *MUC* gene transcript or a MUC polypeptide. The fifth aspect extends to methods for detecting a polymorphism, deletion, mutation, truncation or expansion in a *MUC* gene, a *MUC* gene transcript or a MUC polypeptide, or detecting a level of expression thereof. One embodiment of the fifth aspect is directed to use of an isolated *MUC* nucleic acid to determine whether a mammal has a disease condition, or a predisposition thereto. Another embodiment is directed to use of an isolated MUC polypeptide to determine whether a mammal has a disease condition, or a predisposition thereto.

In a sixth aspect, the present invention provides a method of gene therapy of a disease condition in a mammal, said method including administering to said mammal a gene therapy construct which includes an

isolated *MUC* nucleic acid as hereinbefore defined, to thereby alleviate one or more symptoms of said disease condition in said mammal.

In a seventh aspect, the present invention provides a method of treating a disease condition in a mammal, said method comprising the step of administering to said mammal a pharmaceutically effective amount of a MUC polypeptide or an anti-MUC antibody.

In an eigth aspect, the present invention resides in a pharmaceutical composition comprising a MUC polypeptide or anti-MUC antibody, together with a pharmaceutically acceptable carrier and/or diluent.

Preferably, the mammal is a human.

As used herein, the "disease condition" is associated with aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins.

Preferably, the disease condition is selected from the group consisting of colorectal cancer (CRC), cystic fibrosis (CF), inflammatory bowel disease (IBD), breast cancer (BC), Crohn's disease, ulcerative colitis, asthma and chronic bronchitis.

More preferably, the disease condition is selected from the group consisting of colorectal cancer (CRC), cystic fibrosis (CF), inflammatory bowel disease (IBD) and breast cancer (BC).

As used herein, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1(A): Autoradiograph of a differential display gel showing amplified products from RNA isolated from matched normal colon (N) and primary colorectal tumor (P) tissues. Differentially expressed bands dd29 (*MUC12*) and dd34 (*MUC11*) are arrowed.

FIG. 1(B): Northern blot analysis of total RNA from patient 101 hybridized with the dd29 probe to detect a *MUC12* gene transcript (mRNA). Signal corresponding to 18S ribosomal RNA is shown as a loading control.

FIG. 1(C): Northern blot analysis of RNA from patient 112 hybridized with the dd34 probe to detect a *MUC11* gene transcript (mRNA) . Signal corresponding to 18S ribosomal RNA is shown as a loading control.

FIG. 1(D): Multiplex semi-quantitative RT-PCR showing amplification of *MUC12* mRNA transcripts from matched normal colonic mucosa and primary tumor # 40, normal mucosa from patient # 81 and six colorectal cancer cell lines. Amplification of β_2-microglobulin (β_2-MG) is included as a measure of total RNA.

FIG. 1(E): Multiplex semi-quantitative RT-PCR showing amplification of *MUC11* mRNA transcripts in matched normal colonic mucosa and primary tumors of patients # 40, 164, and 97 and six colorectal cancer cell lines. Amplification of β_2-microglobulin (β_2-MG) is included as a measure of total RNA.

FIG. 1(F): Multiplex semi-quantitative RT-PCR showing amplification of *MUC12* mRNA transcripts from matched normal colonic mucosa and primary tumors # 346, 84, 128, 97 and 316 and from five unpaired Dukes' stage D tumors (M) # 93, 361, 107, 357 and 367. Amplification of β_2-microglobulin (β_2-MG) is included as a measure of total RNA.

FIG. 1(G): Multiplex semi-quantitative RT-PCR showing *MUC11* mRNA transcripts in matched normal colonic mucosa and primary tumors of patients # 110, 346, 84, 128, and 348 and from five unpaired Dukes' stage D tumors (M) # 93, 107, 361, 367 and 357. Amplification of β_2-microglobulin (β_2-MG) is included as a measure of total RNA. Ma denotes molecular size markers in FIG 1D-G.

FIG. 2: Predicted amino acid sequence of *MUC12*. Numbering of amino acids is given on the right. The consensus sequence of the degenerate tandem repeat structure is shown at the top. The two cysteine-rich EGF-like domains are double underlined, a potential coiled-coil domain is in bold, the hydrophobic domain singly underlined and potential N-glycosylation sites shaded. The stop codon is denoted by an asterisk.

FIG. 3: Amino acid sequence alignment of the carboxyl termini of MUC12, hMUC3 (amino acids 1-366), mMuc3 (Shekels *et al.*, 1998, *supra*; amino acids 637-1015), rMuc3 (Gum *et al.*, 1991, *supra*; Khatri *et al.*, 1997, Biochem. Biophys. Acta **1326** 7; amino acids 356-447 and 1-379 respectively), hMUC4 (Moniaux *et al.*, 1998, Biochem. J. **338** 1998; amino acids 861-1156) and rMuc4 (Sheng *et al.*, 1992, J. Biol. Chem. **267** 16341; amino acids 451-744). Light shading demonstrates identity with MUC12 and dark shading highlights all cysteine residues. Hyphens indicate gaps inserted to optimize the alignment.

FIG. 4: Predicted amino acid sequence of *MUC11* showing the degenerate tandem repeat structure. The consensus sequence is shown at the top and amino acids not consistent with this sequence are shown in bold. Hyphens indicate gaps placed in order to optimize the amino acid alignment. A potential N-glycosylation site is shaded.

FIG. 5: mRNA tissue distribution of the 7q22 mucin gene family. Only those tissues showing a positive signal by Northern blot analysis are represented in the histogram. Sixteen tissues of neural origin, heart, aorta, skeletal muscle, bladder, stomach, testis, ovary, spleen, pituitary gland, adrenal gland, thyroid gland, salivary gland and mammary gland were negative for

mucin mRNA expression. Expression was quantified by densitometry and is shown as a proportion of the tissue showing highest expression.

FIG. 6: Domain organization of the C-termini of human MUC12, hMUC3, the rodent Muc3 mucins and the rat and human MUC4 mucins. The relative size of domains is accurate except that the N-glycosylated domain adjacent to the mucin domain in MUC4 is shown at approximately one fifth of its actual size. Only the beginning of the large mucin domains are shown.

FIG. 7: Alignment of the first extracellular EGF-like domain of MUC12 with human EGF-like growth factors. Dark shading highlights identical amino acids and light shading indicates conservative amino acid substitutions.

FIG. 8: Schematic representation of *MUC 11* cloning (A) and *MUC 12* cloning (B).

FIG. 9: Normal colonic expression patterns of MUC11 (A, B) and MUC12 (C) polypeptides as determined by anti-MUC mAb M11.9 and M12.15 immunostaining, respectively. (D) shows MUC 11 gene transcript (mRNA) expression detected by *in situ* hybridization in normal colonic epithelium and loss of expression in CRC (top right).

FIG. 10: Expression of *MUC11* and *MUC12* mRNA in normal colon as detected by RT-PCR. Cytokeratin 20, (CK20) a colonic epithelial marker, was employed as a loading control. 'RC' denotes right colon, 'TC' the transverse colon, 'LC' the left colon, 'SC' sigmoid colon; 'CA' refers to the caecum and 'R' denotes the rectum.

FIG. 11: Expression of *MUC11* and *MUC12* mRNA in CRC cell lines as detected by RT-PCR. The loading control is β_2-microglobulin (B2MG) and 'M' denotes the molecular weight marker.

FIG. 12: Expression of *MUC11* and *MUC12* mRNA in IBD as detected

by RT-PCR. Cytokeratin 20 (CK20) a colonic epithelial marker, was employed as a loading control. 'N' denotes tissues which appear macroscopically normal and 'D' refers to tissues reported to have IBD. 'CA' refers to the caecum, 'CO' the colon, 'LC' the left colon, 'TC' the transverse colon, 'RS' the recto-sigmoid colon, 'SI' the small intestine, 'IL'denotes the ileum and 'IP' an ileal pouch.

FIG. 13: Expression of *MUC11* and *MUC12* mRNA in BC as detected by RT-PCR. The loading control is β_2 microglobulin denoted by B2MG and the molecular weight marker is denoted by 'M'. The positive control was normal colonic cDNA from patient 164.

FIG 14: Northern blot analysis of *MUC11* expression in normal colon (N) and primary CRC (P) of six patients, assessed using a probe corresponding to dd34. The position of ribosomal RNAs are indicated, and signal from 18S ribosomal RNA was used as a loading control.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is predicated in part on the identification of novel *MUC11* and MUC*12* genes which are normally predominantly expressed in the colon. The isolated *MUC* nucleic acids and *MUC* genes of the invention may be useful in treatment and diagnosis of disease conditions associated with aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins. Such disease conditions include but are not limited to cancer of the large bowel (CRC), cystic fibrosis (CF), inflammatory bowel disease (IBD), respiratory diseases such as asthma and chronic bronchitis, breast cancer (BC), ulcerative colitis and Crohn's disease.

The present invention is particularly directed to cancers of the large bowel, which includes the colon, rectum and anal canal, such as CRC, although it extends to biochemically, physiologically and/or genetically related cancers in other parts of the gastrointestinal tract.

The *MUC* genes are, for example, down-regulated in CRC.

By "predominantly expressed" is meant that a *MUC* gene transcript or MUC polypeptide encoded by said *MUC* gene is expressed in the colon at a level greater than in any other organ.

By "associated with" is meant that the disease condition displays symptoms consistent with aberrant Mucin expression, altered properties of mucus or epithelial inflammation involving Mucins. The disease association may be merely correlative or may reflect a causative role of Mucins in the disease condition.

The term "cancer" is used in its broadest sense to include malignant tumours, carcinomas and sarcomas.

In light of the foregoing, it will be appreciated that a *MUC* nucleic acid "corresponds to" a *MUC* gene by being an isolated nucleic acid derived from said *MUC* gene, or a portion thereof. Thus it will be understood that said gene has components including amino acid coding sequences and non-coding sequences. Non-coding sequences include, for example, introns and regulatory sequences which include a promoter, translation initiation and termination sequences and a polyadenylation sequence, for example. The isolated *MUC* nucleic acid may therefore correspond to some or all of the aforementioned components of the corresponding *MUC* gene.

It should be noted that *MUC* terminology has recently undergone revision. In particular, *MUC12* was formerly known as dd 29 or *MUC10.* Also, *MUC11* was formerly known as dd 34. Therefore, with this in mind, should the term *"MUC10"* or "dd29" be encountered herein, it should in all cases be taken to mean *MUC12.*

It will also be understood that a MUC polypeptide is encoded by an isolated *MUC* nucleic acid or by a *MUC* gene as hereinbefore defined.

Isolated *MUC* nucleic acids of the invention may be in DNA (e.g. cDNA or genomic DNA), RNA (e.g. mRNA) or hybrid DNA:RNA form, eithre in double-stranded or single-stranded form. For example, single-stranded *MUC* nucleic acids include nucleic acids having sequences

complementary to the nucleotide sequences of SEQ ID NO:2 and SEQ ID NO:5.

In one embodiment, the isolated *MUC* nucleic acid of the invention comprises a nucleotide sequence having at least 60% identity to the nucleotide sequence according to SEQ ID NO:2, or a nucleotide sequence capable of hybridizing thereto under at least low stringency conditions.

In another embodiment, the isolated *MUC* nucleic acid of the invention comprises a nucleotide sequence having at least 60% identity to the nucleotide sequence according to SEQ ID NO:5, or a nucleotide sequence capable of hybridizing thereto under at least low stringency conditions.

According to these embodiments, it is preferable that the nucleotide sequence has at least 75% identity.

More preferably, the nucleotide sequence has at least 90% sequence identity.

The term "identity" is used herein in its broadest sense to include the number of exact nucleotide or amino acid matches having regard to an appropriate alignment using a standard algorithm, such as but not limited to the Geneworks program (Intelligenetics). For this purpose, BLAST family programs may also be useful (Altschul *et al.*, 1997, Nucl. Acids Res. **25** 3389, which is herein incorporated by reference). A detailed discussion of sequence analysis can be found in Unit 19.3 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Eds Ausubel *et al.*, (John Wiley & Sons), which is herein incorporated by reference.

According to these embodiments, it is preferable that the nucleotide sequence is capable of hybridizing under medium stringency conditions.

More preferably, the nucleotide sequence is capable of hybridizing under high stringency conditions

Reference herein to low stringency conditions includes and

encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridisation at 42°C, and at least about 1 M to at least about 2 M salt for washing at 42°C.

Low stringency conditions also include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridization at 65°C, and (i) 2xSSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 5% SDS for washing at room temperature.

Medium stringency conditions include and encompass from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridisation at 42°C, and at least about 0.5 M to at least about 0.9 M salt for washing at 42°C.

Medium stringency conditions also include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridization at 65°C, and (i) 2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 5% SDS for washing at 42°C.

High stringency includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01 M to at least about 0.15 M salt for hybridisation at 42°C, and at least about 0.01 M to at least about 0.15 M salt for washing at 42°C.

High stringency also includes 1% BSA, 1 mM EDTA, 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS for hybridization at 65°C, and (i) 0.2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 1% SDS for washing at a temperature in excess of 65°C.

In general, washing is carried out at $T_m = 69.3 + 0.41 (G + C)$ % = -12°C. However, the T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatched based pairs.

Although the *MUC* genes and isolated *MUC* nucleic acids of the present invention are exemplified in relation to the human mammalian species, the present invention extends to orthologs in non-human mammals such as in primates, laboratory test animals (e.g. mice, rates, rabbits, guinea pigs, hamsters), companion animals (e.g. dogs, cats), livestock animals (e.g.

sheep, pigs, horses, donkeys, cows) and captive wild animals (e.g. deer, fox).

In light of the foregoing, the term "*MUC* homologs" is used to encompass *MUC* orthologs, isolated nucleic acids which hybridize to *MUC* nucleic acids of the invention and isolated nucleic acids which display at least 60% sequence identity to isolated *MUC* nucleic acids.

It will also be appreciated that *MUC* homologs encompass single or multiple nucleotide substitutions, deletions and/or additions to the isolated *MUC* nucleic acids of the invention, inclusive of mutants, fragments, parts, portions and segments of the nucleotide sequences of the invention.

The isolated *MUC* nucleic acids of the present invention and homologs thereof therefore include oligonucleotides, primers (such as for PCR), antisense sequences, molecules suitable for use in co-suppression and fusion nucleic acid molecules. Ribozymes are also contemplated by the present invention. It will be understood that probes, primers and antisense sequences correspond to distinct portions of isolated *MUC* nucleic acids of the invention, in that they contain nucleotide sequences based on said distinct portions of an isolated *MUC* nucleic acid sequence. Such probe and primer sequences may be based on a *MUC* sequence of the invention by being identical thereto, or by being degenerate with respect thereto.

As used herein, "oligonucleotides" are nucleic acids which comprise a contiguous sequence of no more than seventy (70) nucleotides, whereas "polynucleotides" are nucleic acids which comprise a contiguous sequence of more than seventy (70) nucleotides. A "probe" may be an oligonucleotide or a polynucleotide, either double-stranded or single-stranded, for use in hybridization techniques such as Northern blotting, Southern blotting or *in situ* hybridization. The skilled person will realize that *in situ* hybridization also includes Fluorescence *In Situ* Hybridization (FISH), which is used for determining chromosomal localization. *In situ* hybridization techniques applicable to the present invention will be described in detail hereinafter.

A "primer" is a nucleic acid (usually an oligonucleotide) capable

of annealing to a nucleic acid template under appropriate conditions of ionic strength and temperature, which annealed primer can be extended in a template-dependent fashion by a suitable nucleic acid polymerase (for example *Taq* polymerase or Sequenase™). It will therefore be understood that primers of the invention may be useful for PCR, sequencing, RACE, primer extension and the like.

In use, isolated *MUC* nucleic acids, probes and primers may be modified such as by end-labeling with ^{32}P-ATP and T4 polynucleotide kinase or by random primed labeling with ^{32}P-dCTP and DNA polymerase. Biotinylation is also contemplated, as is modification with phosphorothiorates, fluorochromes, digoxigenin, enzymes and peptides, for example.

It is contemplated that diagnostic methods may be employed which utilize isolated *MUC* nucleic acids of the present invention, or portions thereof such as probes and PCR primers. Also, diagnostic methods employing MUC polypeptides will be discussed in more detail hereinafter.

Diagnostic methods may include detection of *MUC* genes, transcripts and/or polypeptides in samples such as fecal specimens and/or in colonic biopsies, analysis of serum MUC levels in patients with epithelial diseases including cancers, breast tissue biopsy samples or in respiratory mucus samples from patients suffering from CF, asthma or chronic bronchitis.

The diagnostic methods of the present invention may therefore be applicable to determining whether an individual has a disease condition associated with aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins, or a predisposition to said disease. It will be appreciated that "predisposition" as used herein refers to an increased probability that an individual will contract the disease. However, it will also be appreciated that the diagnostic methods may also indicate whether an individual actually suffers from the disease, assist in assessing the severity of disease, a prognosis of the likely course of disease and appropriate treatments for the disease. Thus, the diagnostic methods of the invention may be useful whether or not the individual suffers from one or

more symptoms of the disease.

The present invention therefore contemplates methods of detecting *MUC* genes and *MUC* gene transcripts (e.g. mRNA), such as involving hybridization techniques (for example, by Northern or Southern blotting or *in situ* hybridization) or polynucleotide sequence amplification techniques (for example RT-PCR). Such methods may detect:-

(i) a polymorphism, deletion, mutation, expansion, and/or truncation in *a MUC* gene or *MUC* gene transcript; and

(ii) a relative level of expression of a *MUC* gene transcript (an mRNA transcript derived from a *MUC* gene).

Such methods of detection facilitate determination of whether said *MUC* gene is aberrantly-expressed as an indication of a disease condition or a predisposition thereto. Also, *MUC* gene polymorphisms, deletions, mutations, truncations or deletions may be detected which indicate a disease condition or a predisposition thereto.

It will be appreciated, for example, that measurement of a relative level of expression of a *MUC* gene transcript facilitates diagnostic assessment of whether *MUC* gene expression is downregulated and thereby indicative of CRC.

Although PCR is the preferred nucleic acid sequence amplification technique, It will be appreciated that there are a variety of polynucleotide sequence amplification techniques other than PCR, which include rolling circle amplification (RCA) and strand displacement amplification (SDA). With regard to RCA, reference is made to WO97/19193 which is herein incorporated by reference. With regard to SDA, reference is made to U.S. Patent No. 5455166, which is herein incorporated by reference.

Detailed PCR methods are provided hereinafter, although the skilled person is also referred to Chapter 15 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Eds Ausubel *et al.*, (John Wiley & Sons), which is herein incorporated by reference, for a detailed discussion and examples of PCR methods.

It will also be understood that PCR includes within its scope RT-PCR and multiplex PCR as will be described in detail hereinafter. Such methods may be used for qualitative or semi-quantitative analysis. PCR-based Restriction Fragment Length Polymorphism (PCR-RFLP) methods are also contemplated, which methods are useful when a polymorphism, deletion mutation, truncation and/or expansion either introduces or removes one or more restriction endonuclease sites in a *MUC* gene.

The skilled person will appreciate that Northern, Southern and *in situ* hybridization methods involve formation of a hybrid nucleic acid comprising a *MUC* gene or mRNA transcript and a corresponding isolated *MUC* nucleic acid or portion thereof.

RNA isolation and Northern hybridization methods are described in detail herein, although the skilled person is also referred to Chapter 4 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Eds Ausubel *et al.*, (John Wiley & Sons), which is herein incorporated by reference.

Furthermore, Southern hybridization methods are described in detail herein, although the skilled person is also referred to sections 2.9A-B and 2.10 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Eds Ausubel *et al.*, (John Wiley & Sons), which is herein incorporated by reference.

Also, determining whether a *MUC* gene or *MUC* gene transcript includes a polymorphism, mutation, deletion, truncation and/or expansion can be performed using methods such as PCR-RFLP analysis, Single Strand Conformational Polymorhpism (SSCP) analysis and Denaturing Gradient Gel Electrophoresis (DGGE). These techniques have become well known in the art of mutation detection. A non-limiting example of DGGE is provided in Folde & Loskoot, 1994, Hum. Mut. **3** 83, which is herein incorporated by reference. A non-limiting example of specific allele detection by PCR-RFLP and SSCP is provided in Lappalainen *et al.*, 1995, Genomics **27** 274, which is herein incorporated by reference.

It is proposed that mutations in *MUC11* or *MUC12* genes are associated with bowel cancers (CRC), CF, BC, IBD, chronic bronchitis, asthma, ulcerative colitis and/or Crohn's disease. These are examples of disease conditions associated with aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins. The isolated *MUC* nucleic acids now provide a means for genetic screening of the abovementioned disease conditions in human and other mammalian species. Genetic screening may be conducted by determining full expression or full-length transcript production by Northern blot, cloning and sequencing of the *MUC* genes or identifying mutations by oligonucleotide hybridisation or by direct sequencing of PCR amplification products of the *MUC* genes. In addition, the present invention extends to nucleic acid molecules having translation-terminating mutations leading to truncation mutants. The detection of truncation mutants may be important for genetic analysis of people with, for example, cancer of the large bowel or with a propensity to develop large bowel cancer, determined on, for example, hereditary grounds.

Truncated MUC polypeptides may also be useful in developing therapeutic agents such as antagonists or for developing antibodies. Truncational mutants may be readily detected by a direct protein truncation test. In essence, DNA fragments including PCR amplification products or corresponding mRNA molecules are subjected to *in vitro* translation and optionally also transcription and the translation products assayed by, for example, SDS-PAGE or by differential antibody binding assays. This assay may also be employed to screen for agents capable of inducing truncation mutations or for agents acting as antagonists for truncation mutant-inducing agents.

Alternatively, MUC polypetides may be assayed by, for example, by antibody screening such as in an ELISA.

Thus, it will be appreciated that the present invention contemplates isolated MUC polypeptides, and also:-

(i) polypeptides which comprise an amino acid sequence

having at least 60% identity to a MUC polypeptide amino acid sequence, preferably at least 75% identity thereto, or more preferably at least 90% identity thereto; and

(ii) polypeptides encoded by *MUC* homologs.

Such polypeptides are hereinafter referred to as "MUC homologs".

The MUC polypeptide homologs of the invention include amino acid substitution(s), deletion(s) and/or addition(s) to a MUC polypeptide sequence. Particular examples include antigenic fragments and analogues useful in immunoassays and as therapeutic agents as well as other fragments carrying B cell and/or T cell linear or conformational epitopes. Additions to the amino acid sequence include fusion partners in the form of peptides or polypeptides, which create a MUC fusion polypeptide.

Fusion polypeptides include the MUC polypeptide(s) together with fusion partners such as HIS_6, glutathione-s-transferase (GST), thioredoxin (TR) and maltose binding protein (MBP). Fusion partners greatly assist recombinant synthetic polypeptide purification by virtue of each fusion partner affording affinity purification by a specific affinity matrix. Preferably, the fusion polypeptide also includes a protease-specific cleavage site, so that the fusion partner may be cleaved and removed following purification to leave a substantially unmodified MUC polypeptide.

The use of fusion partners for purification of recombinant expressed polypeptides is well known in the art. Indeed, there are a variety of commercial sources applicable to fusion partners and purification systems such as the QIAexpress™ $(HIS)_6$ system, the Pharmacia GST purification system and the New England Biolabs MBP system.

Also within the scope of fusion partners are "epitope tags". Such tags are well known in the art and include c-myc, influenza hemagglutinin and FLAG tags.

Furthermore, Green Fluorescent Protein (GFP) is a well known

fusion partner applicable to MUC polypeptides of the invention. A particularly useful application of GFP fusion partners is in the visible identification of cells or tissues which express a GFP-MUC fusion polypeptide of the invention. Identification may be performed by flow cytometry or fluorescence microscopy, as are well known in the art.

The MUC polypeptides and MUC homologs of the invention may be in recombinant form of may be chemically synthesized, as is well known in the art. Chemical synthesis is preferably suited to production of MUC peptides. As used herein, "peptides" have no more than fifty (50) contiguous amino acids.

Preferably, MUC polypeptides are in recombinant form.

In order to produce recombinant MUC polypeptides, isolated *MUC* nucleic acids of the present invention may be ligated into an expression vector to form an expression construct capable of directing expression of said *MUC* nucleic acid in a prokaryotic cell (for example, *E. coli*) or in a eukaryotic cell (for example, yeast cells, fungal cells, insect cells, mammalian cells or plant cells).

Suitably, the expression vector comprises one or more regulatory elements which direct expression of the nucleic acid ligated in said expression construct. Such regulatory sequences include promoters, enhancers, splice donor/acceptor sites, polyadenylation sequences, translation initiation (Kozak sequences) and translation termination signals. Suitable promoters may be constitutive (for example, CMV- or SV40-derived promoters) or inducible (for example, Zn responsive metallothionein promoters) or repressible (*tet*-repressible promoters).

Exemplary methods useful for recombinant protein expression and purification, including fusion polypeptides, can be found in Chapters 16 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Eds. Ausubel *et al.*; John Wiley & Sons Inc., 1997 Edition) and Chapters 5 and 6 of CURRENT PROTOCOLS IN PROTEIN SCIENCE (Eds. Coligan *et al.*; John Wiley & Sons Inc., 1997 Edition) which are herein incorporated by reference.

"Analogues" of the MUC polypeptides of the invention contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues. Such chemical analogues may be useful in providing stable means for diagnostic purposes or for producing agonists or antagonists or for producing stable molecules for use in natural product screening.

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with $NaBH_4$.

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with

cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids, contemplated herein is shown in Table 1. Crosslinkers can be used, for example, to stabilise tertiary conformation, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n=1 to n=6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N_α-methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

The present invention further contemplates chemical analogues of the polypeptides of the invention capable of acting as antagonists or agonists thereof, or which can act as functional analogues thereof. Chemical

analogues may not necessarily be derived from the polypeptides of the invention, but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to mimic certain physiochemical properties of MUC poypeptides. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening. Useful sources for screening for natural products include coral, reefs, sea beds, river beds, plants, microorganisms and aqua and antarctic environments.

Still another aspect of the present invention is directed to antibodies specific for MUC polypeptides and/or homologs thereof.

In one embodiment, the anti-MUC antibody is M11.9.

In another embodiment, the anti-MUC antibody is M12.15.

A detailed method of anti-MUC antibody preparation is provided hereinafter.

In this regard, it will be understood that anti-MUC polypeptide antibodies may be produced by immunization with MUC polypeptides or MUC peptides.

In particular, it is also likely that naturally-occurring anti-MUC antibodies may well have naturally arisen against MUC polypeptides.

In light of the foregoing, it will be appreciated that "anti-MUC antibody" as used herein is an antibody specific for, or at least binds to, a MUC polypeptide, irrespective of how the anti-MUC antibody was produced.

The anti-MUC antibodies of the present invention may be useful as therapeutic or diagnostic agents.

For example, a MUC polypeptide or homolog can be used to screen for naturally occurring anti-MUC antibodies. These may occur, for example in some autoimmune diseases. Alternatively, anti-MUC antibodies can be used to screen for MUC polypeptides. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of endogenous MUC polypeptide levels may be important for diagnosis of large bowel cancer or a predisposition to large bowel cancers

or for monitoring certain therapeutic protocols. This knowledge may also be important in other epithelial cancers such as cancer of the breast.

Anti-MUC antibodies of the present invention may be monoclonal or polyclonal. Alternatively, fragments of antibodies may be used such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies. The antibodies of this aspect of the present invention are particularly useful for immunotherapy and may also be used as a diagnostic tool for assessing cancer development or cancer cell apoptosis or monitoring the program of a therapeutic regimum.

For example, anti-MUC antibodies can be used to screen for endogenous MUC polypeptides. The latter would be important, for example, as a means for screening for levels of the MUC polypeptide in a cell extract or other biological fluid or purifying the MUC polypeptide made by recombinant means from culture supernatant fluid. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of the MUC polypeptide.

Both polyclonal and monoclonal antibodies are obtainable by immunization with the enzyme or protein and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal with an effective amount of a MUC polypeptide, or antigenic parts thereof, collecting serum from the

animal, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

The present invention contemplates a method for detecting a MUC polypeptide in a protein extract obtained from a mammal, said method including the step of forming a complex between an anti-MUC antibody and a MUC polypeptide, and then detecting said complex.

The presence of a MUC polypeptide may be determined in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to U.S. Patent Nos. 4,016,043, 4,424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as in the traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized to a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then

added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away and the presence of the antigen is determined by measurement of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In accordance with the present invention the protein extract might be a cell extract, tissue biopsy or possibly serum, saliva, mucosal secretions, lymph, tissue fluid and gastrointestinal fluid. The extract is, therefore, generally a biological sample.

In the typical forward sandwich assay, a first antibody having specificity for MUC or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from 4°C to 37°C) to allow binding of any subunit present in the antibody. Following the incubation period, the solid phase complex is washed and dried and incubated with a second antibody which is specific for a portion of the antigen (i.e. MUC). The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to MUC.

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorochromes or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, such as *via* glutaraldehyde or periodate amongst other means. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody-antigen complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative

visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of antigen which was present in the sample. The term "reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Also, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state of excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to light of the appropriate wavelength and the fluorescence observed indicates the presence of the antigen of interest. Immunofluorescene and EIA techniques are both very well established in the art. Other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

The *MUC* genes of the present invention are likely to function in cell adhesion, signal transduction, growth regulation, epithelial cell protection and/or immunological reactions. The classical gel-forming mucins function in protecting and lubricating epithelial tissues (particularly those of the respiratory and gastrointestinal tracts) by forming a layer of viscoelastic gel. These new mucins, *MUC11* and *MUC12,* show structural similarity to *MUC1. MUC1* can be secreted, but unlike the classical mucins, it is primarily a type I transmembrane protein that interacts and complexes with other adhesion molecules, and is involved in signal transduction. *MUC12* has an EGF growth factor-like domain, is likely to be a transmembrane protein and has a putative tyrosine phosphorylation site that may participate in intracellular signalling. It is hypothesised that loss of *MUC12* may be associated with poor prognosis in CRC.

The isolated *MUC* nucleic acids of the present invention are, therefore, considered in one embodiment, to correspond to cancer suppressor genes. Suppression may mean total inhibition of any development of large bowel cancer or a limitation of the severity of or an amelioration of the condition resulting from a large bowel cancer. The *MUC* nucleic acids of the present invention are also considered in another embodiment to be capable of modulating disease conditions such as CRC, BC, IBD, CF, asthma, chronic bronchitis, ulcerative colitis and/or Crohn's disease

Cystic fibrosis (CF) is an inherited disease of epithelial cell chloride ion transport that affects multiple organ systems. It is the most common cause of severe, progressive lung disease and exocrine pancreatic insufficiency in childhood. The cystic fibrosis transmembrane conductance regulator (CFTR) gene located on chromosome 7q22 encodes a large single chain protein that forms a chloride channel. Virtually all of the morbidity and mortality associated with mutations in the CFTR gene causing cystic fibrosis arise from respiratory disease due to chronic infection and mucus obstruction. The precise mechanism of mucus accumulation in cystic fibrosis is controversial. Data suggest that CFTR malfunction may trigger mucin secretion and alter mucus properties, and/or bacterial infection triggers the hypersecretion of mucin in CF patients. The gene of the present invention is expressed in the colon, pancreas, small intestine, and lung, all tissues where mucus obstruction occurs. Accordingly, aberrant expression of the genes may contribute to cystic fibrosis.

Aberrant mucin expression is also a recognised component of IBD. Inflammatory bowel disease is characterised by considerable alterations in glycosylation, sialyation and sulphation of glycoproteins. It is unclear whether the changes in mucus production are a cause or response to the disease. Susceptibility genes for inflammatory bowel disease have been localised to chromosomes 3, 12 and 7q22. Accordingly, the *MUC* genes of the present invention are considered candidates for susceptibility genes for

IBD. Up or down regulation, or altered secretion of one of these mucins may influence the quality of colonic mucus and therefore the pathology of these diseases. Certain inherited forms of these genes may indicate a predisposition to IBD.

The identification of *MUC* genes and isolated *MUC* nucleic acids permits the generation of a range of therapeutic methods and compositions. Such therapeutics may modulate *MUC* gene expression and the activity of MUC polypeptides. Modulators contemplated by the present invention includes agonists and antagonists of *MUC* gene expression. Antagonists of *MUC* gene expression include antisense molecules, ribozymes and co-suppression molecules. Agonists include molecules which increase promoter activity or interfere with negative mechanisms. Agonists of *MUC* include molecules which overcome any negative regulatory mechanism. Antagonists of MUC polypeptides include antibodies and inhibitor peptide fragments. Another class of therapeutics may be designed to mimic or block intracellular signal transduction by MUC polypeptides.

In accordance with the present invention, it is proposed that MUC functions as a suppressor of cancer development in the large bowel. Hereditary cancers arise with loss of the wild-type gene. In addition, germline mutations underlying large bowel cancer are inactivated for the *MUC* genes and, therefore, hereditary cancers have no functional copy of the gene. Furthermore, sporadic large bowel cancers arise with somatic loss of both copies of the gene. The present invention extends to the use of modulating levels of expression of *MUC* genes or their translation products in the context of cancers related thereto.

Thus, the present invention contemplates a method of gene therapy of a mammal. Such a method utilizes a gene therapy construct which includes an isolated *MUC* nucleic acid ligated into a gene therapy vector which provides one or more regulatory sequences that direct expression of said nucleic acid in said mammal.

Such regulatory sequences may include a promoter, an

enhancer, a polyadenylation sequence, splice donor/acceptor sequences and translation termination and intiation sequences.

Typically, gene therapy vectors are derived from viral DNA sequences such as adenovirus, adeno-associated viruses, herpes-simplex viruses and retroviruses. Suitable gene therapy vectors currently available to the skilled person may be found in Robbins *et al.*, 1998, Trends Biotechnol. **16** 35, for example, which is herein incorporated by reference.

If "anti-sense" therapy is contemplated, then one or more selected portions of a *MUC* nucleic acid may be oriented 3'→5' in the gene therapy vector.

Administration of the gene therapy construct to said mammal, preferably a human, may include delivery via direct oral intake, systemic injection, or delivery to selected tissue(s) or cells, or indirectly via delivery to cells isolated from the mammal or a compatible donor. An example of the latter approach would be stem-cell therapy, wherein isolated stem cells having potential for growth and differentiation are transfected with the vector comprising the *MUC* nucleic acid. The stem-cells are cultured for a period and then transferred to the mammal being treated.

Delivery of said gene therapy construct to cells or tissues of said mammal or said compatible donor may be facilitated by microprojectile bombardment, liposome mediated transfection (e.g. lipofectin or lipofectamine), electroporation, calcium phosphate or DEAE-dextran-mediated transfection, for example. A discussion of suitable delivery methods may be found in Chapter 9 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Eds. Ausubel *et al.*; John Wiley & Sons Inc., 1997 Edition), for example, which is herein incorporated by reference.

For example, a *MUC* nucleic acid may be introduced into a cell to enhance the ability of that cell to survive, conversely, *MUC* antisense sequences such as 3'→ 5' oligonucleotides may be introduced to decrease the survival capacity of any cell expressing an endogenous *MUC* gene.

In this regard, increased *MUC* expression or activity is

important in conditions of repressing cancer growth and/or development. Decreased *MUC* expression or activity may be important, for example, in the treatment of cystic fibrosis or the treatment of inflammatory bowel disease.

Accordingly, the present invention contemplates a pharmaceutical composition comprising a MUC polypeptide or a derivative thereof or a modulator of *MUC* gene expression or activity, inclusive of anti-MUC antibodies. These components are referred to herein as the "active ingredients", and are suitably provided in combination with one or more pharmaceutically-acceptable carriers and/or diluents.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) and sterile powders for the extemporaneous preparation of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or, for example, water, ethanol, polyol (for example, glycerol, propylene glycol and liquid polyethylene glycol, and the like) or suitable mixtures thereof as well as vegetable oils. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thirmersal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously

sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 µg and 2000 mg of active ingredient.

The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter: a binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the

amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the treatment of disease in living subjects having a diseased condition in which bodily health is impaired.

The principal active ingredient is compounded for convenient and effective administration in effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 µg to about 2000 mg. Expressed in proportions, the active compound is generally present in from about 0.5 µg to about 2000 mg/mL of carrier. In the case of compositions containing supplementary active ingredients, the dosages are determined by reference to the usual dose and manner of administration of the said ingredients. It is

also convenient to represent the effective amounts of active ingredients as an amount per kg body weight. For example, the present invention encompasses effective amounts for 0.005 µg/kg body weight at 2000 mg/kg body weight.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating *MUC* gene expression or MUC polypeptide activity. The vector may, for example, be a viral vector.

From the foregoing, it is apparant that therapeutic methods and compositions of the invention are useful in the treatment of disease conditions associated with aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins.

Preferably, the disease condition is selected from the group consisting of colorectal cancer (CRC), cystic fibrosis (CF), inflammatory bowel disease (IBD), breast cancer (BC), Crohn's disease, ulcerative colitis, asthma and chronic bronchitis.

More preferably, the disease condition is selected from the group consisting of colorectal cancer (CRC), cystic fibrosis (CF), inflammatory bowel disease (IBD) and breast cancer (BC). although not limited thereto. The therapeutic methods of the invention may therefore be used to alleviate one or more symptoms of diseases or be used as prophylactic treatments to prevent, or reduce the likelihood of, said symptoms from occurring.

The present invention is further described by the following non-limiting Examples.

EXAMPLES

EXAMPLE 1: ***Tissue Specimens***

Tissue specimens were collected from patients undergoing surgery (Dukes' A n=5; Dukes' B n=5, Dukes C n=5, Dukes' D n=5). Colonic specimens were obtained from patients undergoing either colectomy or

partial hepatectomy for colorectal carcinoma. Samples of normal colonic mucosa, primary colon cancer, liver metastases (if present) and adjacent normal liver were rapidly excised from operative specimens, snap-frozen in liquid nitrogen and stored at -70°C until use. Care was taken to exclude normal mucosal tissue from tumour samples. Junctional tissue specimens from four tumours of each Dukes' stage were randomly selected for *in situ* hybridisation. Tissues were fixed for 24-48 hours in 10% v/v buffered formalin, dehydrated in ethanol, cleaned in chloroform and embedded in parraffin wax. Biopsy specimens of normal colonic epithelium from four distinct regions of the colon were collected via colonoscopy from each of three healthy individuals undergoing routine colonoscopic screening. Similarly, intestinal biopsies were obtained via colonoscopy from ten patients with inflammatory bowel disease. Specimens were snap-frozen and stored at -70°C until RNA was extracted as per Example 3 below.

EXAMPLE 2: ***Cell Lines and Culture***

Seven human colonic tumour lines were obtained: LIM1215, LIM2405, LIM1863, LIM1899 (Ludwig Institute, Melbourne, Australia), HT29 (ATCC HTB38), SW480 (ATCC CCL 228) and SW620 (ATCC CCL 227). LIM1215 and SW620 are each derived from CRC metastases. Cell lines were maintained in RPMI 1640 with 10% v/v fetal calf serum, 2 mM glutamate, 25 mM HEPES, 60 mg/ml penicillin G and 100 mg/ml streptomycin sulfate and incubated in 5% v/v CO_2 and 95% v/v air at 37°C. Cultures were passaged twice weekly using standard techniques. The following breast carcinoma lines were included in this study: KPL-1 (a gift of Dr Junichi Kurebayashi, Suzuki, Japan), MA11 (a gift of Dr Philip Rye, Oslo, Norway), BT 20, DU4475, MCF-7, MDA-MB-453, SK-Br-3, T47D, UACC-893, ZR-75-1 and ZR-75-30 (ATCC, Rockville, MD), and MDA-MB-435 and MDA-MB-468 (a gift of Dr. Janet Price, MD Anderson Cancer Center, Houston, TX). All breast cancer cell cultures were maintained in RPMI-1640 medium supplemented with 10% fetal calf serum and 0.006% penicillin and 0.01% streptomycin with the following exceptions: DU-4475 in RPMI-1640 with 20% FCS, KPL-1 was

maintained in DMEM with 5% FCS, MA11 in 1:1 Ham's F12:RPMI-1640 with 10% FCS, SK-Br-3 in McCoy's medium with 15% FCS, and UACC-893 in RPMI-1640 with 15% FCS.

EXAMPLE 3: ***RNA Extraction***

Total RNA was isolated by the method of Chomczynski and Sacchi (Chomczynski *et al.*, 1987, Anal. Biochem. **162** 156). Cells were resuspended in RNA extraction buffer (4 M guanidinium isothiocyanate containing 25 mM sodium citrate, pH7.0, 0.5 % w/v sodium lauroyl sarcosine (SLS) and 0.1 M 2-mercaptoethanol). Tissue samples were homogenised in RNA extraction buffer. Extracted RNA was dissolved in RNase-free water and the concentration and purity determined by spectrophotometry at 260 and 280 nm (Sambrook *et al.*, Molecular Cloning, A Laboratory Manual. 2nd Ed. Cold Spring Harbour Laboratory Press. Cold Spring Harbour, NY, 1989). The integrity of the RNA was assessed by denaturing agarose gel electrophoresis and samples transferred to HYBOND N (Amersham, Bucks, England) membrane by capillary blotting.

EXAMPLE 4: ***DNA Sequencing***

Approximately 500 ng of DNA were employed in a cycle sequencing reaction with 2.5 pmol of primer and 4 µl of Dye terminator or dRhodamine reaction mix (DNA Cycle Sequencing Kits, Perkin Elmer, Norwalk, CT,) in a total volume of 10 µl. Reaction mixes contained *Amplitaq* DNA polymerase, dNTPs and fluorescently labelled dideoxynucleotides (dye terminators). Cycling reactions were as follows: 25 cycles of denaturation at 96°C (30 s), primer annealing at 50°C (15 s) and extension at 60°C (4 min). Unincorporated nucleotides were removed by ethanol precipitation. The reactions were analysed on a Model 373A automated DNA sequencer (Applied Biosystems) run by technical staff in the core sequencing facility of the Queensland Institute of Medical Research.

EXAMPLE 5: ***Identification by Differential Display of Two cDNAs Encoding Mucins Downregulated in Colorectal cancer***

The differential display method was devised from the original technique described by Liang & Pardee, 1992, Science **257** 967. Total RNA was isolated by the method as described previously. Reverse transcription was carried out using one of four anchored primers, $T_{12}MG$, $T_{12}MC$, $T_{12}MA$ and $T_{12}MT$ (Operon Technologies Inc., Alameda, CA) and Superscript RNAse H- reverse transcriptase (Gibco BRL, Gaithersburg, MD). One arbitrary 10mer primer (Operon Technologies Inc.) was selected at random to be employed in a PCR with the appropriate anchored primer. Two patients, 101 and 112, were analyzed simultaneously and duplicates of two separate reverse transcription reactions electrophoresed on each gel. Gels were put down wet and autoradiographed for 1-3 days. DNA was removed from gel slices by boiling and reamplified by PCR. Bands were then cloned into pGEM-T (Promega Corporation, Madison, WI) and sequenced. Sequences were analysed by multiple sequence similarity searches using BLAST algorithms (Altshcul *et al.*, 1990, *supra*) accessed through the National Centre of Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov).

Differential display was performed on RNA from paired normal colonic mucosa and primary colorectal cancers. Using a PCR primer combination of $T_{12}MG$ and 10mer 5'-ACTTCGCCAC-3' (SEQ ID NO:7), bands dd29 (*MUC12*) and dd34 (*MUC11*) were both amplified from normal colonic mucosal RNA of two patients and were consistently downregulated in the tumors from these patients in multiple PCR reactions (FIG. 1A). Following reamplification PCR, discrete bands of approximately 720 bp for dd29 and 530 bp for dd34 were isolated and cloned into pGEM-T. Sequence analysis showed that both cDNAs were novel, with no match in any database accessed through the NCBI. Repetitive segments typical of mucin tandem repeats were observed in dd34.

EXAMPLE 6: ***Northern Blot Analysis***

Northern blot analysis was performed on paired normal and tumor total RNA extracted from the same patients employed in the differential display experiment. dd29 (*MUC12*) and dd34 (*MUC11*) were random primer-

labeled using a Megaprime DNA labeling system (Amersham, Aylesbury, UK) and hybridization performed at 65°C in buffer containing 7% SDS, 0.26 M Na_2HPO_4, 1 mM EDTA, 1% BSA.

Northern blot analyses of dd29 (FIG. 1B) and dd34 (FIG. 1C) with colonic total RNA used for the differential display reactions revealed a polydisperse signal beginning near the top of the gel for RNA isolated from normal colonic mucosa and no signal in tumor-derived RNA. Probe dd29 showed some cross-hybridization to ribosomal RNA. Polydispersity of signal is a hallmark of mucin RNA blots due to shearing of very high molecular weight transcripts.

EXAMPLE 7: ***Multiplex Semi-quantitative RT-PCR***

Multiplex semi-quantitative RT-PCR was performed on total RNA isolated from six colorectal cancer cell lines and from paired normal colonic mucosa and tumor colorectal cancer tissues from 20 patients, five of each Dukes' stage. Informed consent was obtained from each subject after approval by the appropriate hospital Ethics Committee. PCR products were quantitated relative to a β_2-microglobulin cDNA amplification control using densitometry. First strand cDNA synthesis was accomplished using 1 µg of total RNA. PCR amplification of cDNA was performed in a total volume of 25 µl containing 1 µl of the first strand cDNA synthesis reaction products, 2.5 µl 10x *Taq* polymerase buffer (25 mM TAPS (tris-[hydroxymethyl]-methyl-amino-propane-sulfonic acid, sodium salt) pH 9.3, 50 mM KCl), 2 mM dNTPs, 25 mM $MgCl_2$, 20 pmol each of the forward and reverse primers, and 2.5 U *Taq* polymerase. Gene-specific forward and reverse primers for *MUC12* and *MUC11* were designed to produce PCR products of 510 bp and 169 bp respectively. Primers for β_2-microglobulin generated a PCR product of 247 bp (Gussow *et al.*, 1987, J. Immunol. **139** 3132). Primers were:

MUC12F1; 5'-TGAAGGGCGACAATCTTCCTC-3' (SEQ ID NO:8);

MUC12R1; 5'-TACACGAGGCTCTTGGCGATGTTG-3' (SEQ ID NO:9);

MUC11F1; 5'-CAGGCGTCAGTCAGGAATCTACAG-3' (SEQ ID NO:10);

MUC11R1; 5'-GAGGCTGTGGTGTTGTCAGGTAAG-3' (SEQ ID NO:11);

β-21F; 5'-TGAATTGCTATGTGTCTGGGT-3' (SEQ ID NO:12);

β-21R; 5'-CCTCCATGATGCTGCTTACAT-3' (SEQ ID NO:13);

MUC12TOTF1;5'-AGCCAACCAGGCTCAGCTCT-3' (SEQ ID NO:14); and

MUC12TOTR1;5'-GCTCACACAGTGGATGCTACC-3' (SEQ ID NO:15).

After an initial denaturation step of 94°C for 5 minutes, the amplification conditions were: 21 cycles of denaturation at 94°C (30 s) for MUC12, (24 cycles of denaturation at 94°C (30 s) for MUC11), annealing at 60°C (30 s) and extension at 72°C (30 s). PCR products were electrophoresed on 1.2% 1x TBE gels and photographed.

Due to the polydisperse signals obtained by Northern analysis, expression of *MUC11* and *MUC12* was examined in a range of colorectal cancer cell lines and tissue mRNAs by multiplex semi-quantitative RT-PCR. dd29 was not expressed in any of six colorectal cancer cell lines examined (FIG. 1D). In contrast, dd34 showed a different pattern of expression, with HT29, LIM1215, LIM1899, LIM1863 lines revealing very faint PCR products, and SW620 and SW480 lines showing relatively high levels of expression (FIG.1E). For tumor tissue-derived RNA, downregulation was defined as amplified band intensity less than 30% of that observed from paired normal colon tissue. dd29 was found to be downregulated or absent in 6/15 (40%) tumors with paired normal samples, and at low levels in 3/5 (60%) Dukes' stage D samples (where normal colon was not available for comparison) (FIG.1F). dd34 was downregulated in the tumors of 12/15 (80%) paired samples and expressed at low levels in 4/5 (80%) Dukes' stage D samples. One of five Dukes' stage D samples showed relatively high levels of expression of dd34 (FIG. 1G). Significantly, 13/15 (87%) colorectal cancers showed downregulation of at least one of these mucin genes, with 5/15 (33%) showing downregulation of both genes.

EXAMPLE 8: ***Differential Tissue Distribution of MUC11, MUC12 and MUC3 mRNAs***

A human RNA "master blot" (Clontech, Palo Alto, CA, catalogue number 7770-1) with RNA from 50 different tissues and controls was used to

examine mucin gene expression. DNA fragments encoding dd29, dd34 and *MUC3* (Genbank Accession No. M55405, a gift from Dr. Sandra Gendler, Mayo Clinic, Scotsdale, Arizona) were excised from vector and radiolabeled as described above. Hybridization was performed as per the manufacturer's instructions. The master blot was reprobed with a radiolabeled β-actin cDNA as a loading control.

Analysis of the tissue distribution of *MUC11*, *MUC12* and *MUC3* transcripts in RNA isolated from 50 different normal tissues showed a distinct pattern of expression for each gene (FIG. 5). *MUC12* and *MUC11* showed highest expression in colon but had different patterns in other organs, mainly restricted to those of epithelial type. *MUC11* had a wider epithelial distribution than *MUC12* which was restricted to expression in the colon, and weakly in the pancreas, prostate and uterus. Consistent with published findings (Van Klinken *et al.*, 1997, Biochem. Biophys. Res. Comm. **238** 143), *MUC3* was found to be predominantly expressed in the small intestine and at much lower levels in the colon. Interestingly, it was also present in the thymus.

EXAMPLE 9: ***Extending the Sequences of dd29 and dd34***

The strategy employed in the cloning of *MUC11* and *MUC12* is shown in FIGS. 8A and 8B respectively.

9.1 *Library Screening*.

A λgt11 human fetal brain 5'-STRETCH PLUS cDNA library (Clontech, Palo Alto, CA) was screened using radiolabeled dd29 and dd34. λ DNA was extracted and inserts were excised, cloned into pBSK- and sequenced.

9.2 *PCR to extend the sequence of dd34 by linking clones 2 and li5*

Screening of the fetal brain library with clone dd34, yielded two new cDNA clones: clone 2 (1043 bp) and clone li5 (1045 bp). Clone dd34 was a perfect match to the middle of the larger clone 2. cDNA from clone li5, however, was highly homologous but not identical to the cDNA from clone dd34. To ascertain whether these partial cDNAs arose from a single mRNA

transcript, RT-PCR was carried out using combinations of forward and reverse primers specific for each cDNA in an attempt to link them. RT-PCR was performed on total RNA extracted from normal colon in a stringent touchdown PCR using high fidelity DyNAzyme DNA polymerase (Finnzymes, Espoo, Finland). Primer combination MUC11F1 and li5R (5'-GGGAACACTGTGGTTTCAGTTGAG-3'; SEQ ID NO:16) yielded a PCR product of 2 kb demonstrating that these two cDNAs were derived from a single transcript. This product was cloned into pGEM-T and sequenced.

9.3 *PCR library screening - to extend sequence of dd29*

Forward and reverse primers for dd29 (dd29F1 and dd29R1) were used in combination with a T7 vector-derived primer in a stringent touchdown PCR to screen an ulcerative colitis (UC) plasmid library (a gift from Dr. Jonathon Fawcett, Queensland Institute of Medical Research, Brisbane, Australia). Amplified products were purified, cloned into pGEM-T and sequenced.

EXAMPLE 10: ***Sequence Analysis of dd29 (MUC12)***

The sequence of dd29 revealed that it was amplified as a result of priming of random 10mer at both ends of the PCR product and that it did not contain a 3' untranslated region (3'-UTR) or poly A tail. Screening of an UC cDNA library with dd29-specific primers extended the sequence 840 bp in the 5' direction and 800 bp in the 3' direction to the poly A tail (Genbank Accession Number AF147790). To confirm contiguous cDNA sequence, primers MUC12TOTF1 and MUC12TOTR1 were designed to produce an expected PCR product of 1532 bp; primers corresponded to bases 230-250 and 1742-1762, respectively, in SEQ ID NO:6. In a stringent touchdown PCR amplification procedure an intense discrete product of the expected size was identified from normal colonic cDNA and cDNA from the Caco-2 colonic cancer cell line. This reaction confirmed the reported *MUC12* cDNA sequence.

Conceptual translation of the composite *MUC12* cDNA reveals the presence of serine/threonine and proline-rich degenerate tandem repeats

(FIG. 2) consistent with this protein being a member of the epithelial mucin family. The deduced 28 amino acid tandem repeat structure is shown in FIG. 2. Following the mucin-repeat domain, MUC12 contains two cysteine-rich EGF-like domains separated by a 150 amino acid non-mucin-like sequence (amino acids 261-410) containing five N-glycosylation sites and a potential coiled-coil domain. The second cysteine-rich EGF-like domain is immediately followed by a putative transmembrane domain containing 26 hydrophobic or uncharged amino acids, and a cytoplasmic tail of 75 amino acids at the carboxyl terminus.

Sequence alignment of MUC12, human MUC3 (hMUC3), rat Muc3 (rMuc3), mouse Muc3 (mMuc3), human MUC4 (hMUC4) and rMuc4 is shown in FIG. 3. When aligned by the transmembrane amino acid sequences, MUC12 was found to have areas of significant homology to rMuc3, mMuc3 and hMUC3, including perfect conservation of eight cysteine residues in the second EGF-like domain. With inclusion of three small gaps, each of these cysteines also align with those in rat and human MUC4. Interestingly, all six mucins contain a conserved EGF-like sequence of Cx(5)GPxCxCx(9)GExC. Furthermore, there is some (4 out of 8) conservation of the cysteine residues between MUC12 and the human and rodent MUC3 and MUC4 mucins in the first EGF-like domain.

EXAMPLE 11: ***Sequence Analysis of dd34 (MUC11)***

Clone dd34 (544 bp) was also obtained as a result of priming of random 10mers at both ends of the PCR product. Screening of a λgt11 human fetal brain library yielded two positive plaques which hybridized to dd34, clone li5 (1045 bp) and clone 2 (1043 bp). These two clones represented opposite ends of a 2.8 kb partial *MUC11* cDNA sequence (Genbank Accession Number AF147791), the linking of which was established by PCR (see Methods). Conceptual translation of the *MUC11* composite is shown in FIG. 4. The entire 957 amino acid sequence consisted of serine, threonine and proline-rich tandem repeats of 28 amino acids in length, consistent with it being derived from a large epithelial mucin. The

deduced tandem repeat structure and consensus repeat sequence for MUC11 is shown in FIG. 4.

EXAMPLE 12: ***Chromosomal Localization of MUC11 and MUC12***

DNA fragments excised from dd29 (720 bp) and dd34 (530 bp) were nick translated with biotin-14-dATP and hybridized *in situ* at a final concentration of 10 ng/µl to metaphases from two normal males. The fluorescence *in situ* hybridization (FISH) method was modified from that previously described (Callen *et al.*, 1990, Ann. Genet. **33** 219) in that chromosomes were stained before analysis with both propidium iodide as counterstain and DAPI for chromosome identification. Images of metaphase preparations were captured by a cooled CCD camera using the CyroVision Ultra image collection and enhancement system (Applied Imaging Int Ltd, Sunderland, U.K.).

Twenty metaphases from a normal male were examined for hybridization to dd29 and dd34 probes. For both genes, all of the metaphases showed strong signal on one or both chromatids of chromosome 7, at band 7q22 (data not shown). A similar result was obtained using metaphases from a second normal male.

EXAMPLE 13: ***Production of monoclonal antibodies reactive with MUC11 and 12***

The following peptides were conjugated to keyhole limpet haemocyanin (KLH) with the heterobifunctional cross-linking agent m-maleimidobenzoyl-N-hydroxysuccinimide ester using standard techniques (Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor, Cold Spring Harbor Laboratory, 1988, which is herein incorporated by reference):-

MUC11: CFHSRPASTHTTLFTED (SEQ ID NO: 17); corresponding to part of the degenerate tandem repeat region, specifically amino acid residues 690-705 deduced from the partial cDNA MUC11 clone, with an N-terminal cysteine residue added for conjugation);

MUC12: TYRNFTEKMNDASSQEC (SEQ ID NO: 18); corresponding to part of the N- glycosylated region, specifically amino acid residues 286-302 deduced from the partial cDNA MUC12 clone, with a C-terminal cysteine residue added for conjugation).

One Balb/c mouse was immunised with each KLH-conjugated peptide as per the following protocol:-

Day 0: KLH-conjugated peptide was diluted to 100 µg/mL in phosphate buffered saline (PBS) and mixed with an equal volume of complete Freund's adjuvant (CFA). Each mouse was injected intra-peritonealy with 0.5 mL of this mixture.

Day 14: Each mouse was immunised as above but peptide was mixed with incomplete Freund's adjuvant (IFA).

Day 33: Each mouse was immunised as on day 14.

Day 43: Each mouse was bled from the tail to assess antibody production by ELISA (see below).

Day 53: Each mouse was injected intra-venously with 100 µL of peptide at 100 µg/mL in PBS without adjuvant, and with 100 µL mixed with IFA intra-peritonealy.

Day 56: Mice were euthanased, and the spleen removed for fusion with myeloma cells.

Splenocytes were fused to Ag8 mouse myeloma cells at a ratio of 5:1 with polyethylene glycol using established methods (Harlow & Lane, *supra*).

Specific antibody producing clones were screened by a solid phase antigen antibody capture ELISA with the immunizing peptides bound to polystyrene assay plates using established methods (Harlow & Lane, *supra*). Positive clones were expanded, retested for specific antibody production and recloned by limiting dilution. Clones were further tested for reactivity with paraffin embedded normal colonic mucosa.

13.1 *MUC11 and MUC12 reactive hybridomas*

Two hybridomas, one reacting with each of MUC11 and MUC12 peptides and with paraffin embedded colonic sections are described in Table 2.

13.2 *Immunohistochemical detection of MUC11 and MUC12 in normal colonic epithelium using antibodies M11.9 and M12.15*

Paraffin sections (4 µm) of normal colonic epithelium were dewaxed with xylene, hydrated in a graded series of ethanol to water and treated with 0.1 U/mL neuraminidase (Boehringer, Germany) in 50 mM Na acetate, 150 mM NaCl, 100 mM $CaCl_2$ buffer, pH 5.5 for 1 hr at room termperature to remove sialic acid groups. Sections were then treated with 1% H_2O_2, 0.1% NaN_3 in Tris buffered saline (TBS) for 10 min to quench endogenous peroxidase activity, and non-specific protein binding blocked with 4% skim milk in TBS for 15 min. Monoclonal antibodies M11.9 and M12.15 were semi-purified by PEG precipitation and diluted to 5-50 µg/mL in TBS/50% non-immune goat serum and incubated for 2 hours overnight at room temperature. Sections were washed once with 1% TX-100 in TBS for 5 min and then twice in TBS for 5 min. Sections were incubated for 30 min at room temperature with pre-diluted biotinylated goat anti-mouse immunoglobulins (Zymed, USA) and then washed as above. Sections were then incubated for 15 min at room temperature with pre-diluted streptavidin-conjugated horseradish peroxidase (Zymed Laboratories) and then washed as above. Peroxidase activity was detected using 10 mg/mL 0.05% diaminobenzidine, 0.03% H_2O_2 in Tris saline, pH 7.6. Sections were counterstained with haematoxylin, dehydrated with ethanol, cleared with xylene and mounted in DePeX.

M11.9 reacts strongly with colonic epithelium, primarily with columnar cells of the surface epithelium (see FIG. 9A). Both goblet and columnar cells deep in the crypts are not stained by this antibody (see FIG. 9A). In surface epithelial columnar cells M11.9 reacted with the perinuclear cytoplasm, lateral cell membranes and most strongly as granular staining in the subapical cytoplasm (FIG. 9B). This localisation suggests reactivity with

precursor in the rough endoplasmic reticulum (perinuclear staining), reactivity with mature mucin on the lateral membranes at columnar cell junctions with other cells, and reactivity with processed mature mucin in granules for apical secretion or incorporation into the apical cell membrane. This pattern of reactivity is distinct from that seen for other known mucin core proteins.

M12.15 also reacts strongly with colonic epithelium, and like M11.9 it reacts primarily with columnar cells of the surface epithelium (see FIG. 9C). However, M12.15 gave a more diffuse cytoplasmic staining pattern than that seen with M11.9, although, like M11.9, the strongest staining was in the apical cytoplasm.

Immunohistochemistry in normal colonic mucosa with these antibodies demonstrates protein expression of the MUC11 and MUC12 gene, supporting the mRNA studies. The co-expression of MUC11 and MUC12 in normal colon is also consistent with the RT-PCR data showing similar levels of relative expression of these two mucin genes in different regions of the intestinal tract.

EXAMPLE 14: Expression of *MUC11* by *in situ* hybridization

14.1 *Methods*

Optimisation of conditions for *in situ* hybridisation, outlined below, was based upon published techniques (Rex & Scotting, 1994, Biochemica **3** 24, which is herein incorporated by reference). Riboprobes were made by *in vitro* transcription of DNA with SP6 and T7 RNA polymerases and incorporation of a digoxigenin-labelled uridine triphosphate (DIG-UTP). The orientation of inserts in pGEM-T was established by sequencing. Insert in the antisense direction and thus complementary to RNA template was the hybridisation probe and insert in the sense direction was used as a negative control. 1 mg of purified linearised plasmid pGEM-T was labelled in the presence of 1/10 volume 10 x transcription buffer, 1/10 volume 10 x NTP mix (1 mM ATP, CTP, GTP, 0.65 mM UTP, 0.35 mM DIG-UTP), 10U RNase inhibitor and 40U of either SP6 or T7 RNA polymerase. The reaction was carried out at 37°C for 2 hours and terminated by addition of 2

µl of 0.2M EDTA. Probes were ethanol precipitated with 1/11 volume 4M LiCl and placed at -20°C for 2 hours. They were then centrifuged at 12,000 g for 30 min at 4°C. Pellets were washed with 70% ethanol, air-dried for 10 min and resuspended in 100ml of RNase-free water.

Paraffin-embedded junctional tissue specimens were sectioned at 4 µm onto sterile water and affixed to Vectabond-treated slides (Vector Laboratories). Sections were dewaxed in xylene, rehydrated and then incubated for 5 min in 0.2 N HCl. HCl treatment contributes to an improvement in the signal to noise ratio by extraction of proteins and partial hydrolysis of target sequences. Slides were washed in sterile water for 5 min, followed by 5 min in PBT (PBS and 0.1% Tween 20). Sections were then incubated in proteinase K (5 mg/ml) at 37°C for 15 min and washed briefly in 3 x PBT. They were fixed in 4% paraformaldehyde for exactly 20 min and prehybridised for 4 hours at 70°C in hybridisation buffer (50% formamide, 5 x SSC, 1% SDS, 500 mg/mL tRNA, 50 mg/mL heparin). Denatured probe (0.5 mg/section) was added to hybridation buffer and sections hybridised overnight at 70°C.

Sections were washed in 2 x wash solution 1 (50% formamide, 5 x SSC, 1% SDS) at 65°C followed by 2 x washes in wash solution 2 (50% formamide, 2 x SSC) also at 65°C. Sections were then incubated with anti-digoxygenin-AP antibody at 1/2000 in PBS overnight at 40°C.

Excess antibody was removed by 3 x 20 minute washes in PBT. Sections were then washed 2 x 20 minute in NTMT buffer (100 mM Tris, (pH 9.5), 50 mM $MgCl_2$, 100 mM NaCl, 0.1% Tween 20, 2 mM levamisole). Hybridisation was visualised with NBT and BCIP overnight at room temperature. The reaction was stopped by immersion of slides into 1 x TE and sections lightly counterstained in eosin. Sections were then dehydrated through ethanols of increasing concentration to xylene and mounted in DePeX. Slides were photographed within 3 days due to fading of the signal with time.

14.2 *Detection of MUC11 mRNA by in situ hybridization*

Intense signal for *MUC11* was observed in the columnar cells of the surface epithelium in all specimens of the normal colon. However, it was not possible to conclusively identify positive signal in the goblet cells of the colonic epithelium. Transcripts for *MUC11* were not detected in adjacent carcinoma of several junctional tissue specimens (an example is shown in FIG. 9D), thus confirming the findings of the differential display and Northern blot analyses.

EXAMPLE 15: ***Expression of MUC11 and MUC12 in normal colon by RT-PCR***

The results of RT-PCR experiments to determine the expression patterns of *MUC11* and *MUC12* genes in normal colonic epithelium are shown FIG. 10.

MUC11 and *MUC12* are predominantly expressed in the colon, although the data in FIG. 10 show that in fact their levels of expression vary within the colon. In this regard, a progressive increase (3-4 fold) in the expression of both *MUC11* and *MUC12* was seen from the right colon to the rectum.

EXAMPLE 16: ***Expression of MUC11 and MUC12 in CRC by RT-PCR and Northern hybridization***

The expression patterns of *MUC11* and *MUC12* in CRC were investigated by RT-PCR, and the results are shown in FIG. 11. After 40 rounds of amplification, *MUC11* expression was observed in all CRC cell lines under investigation. Similarly, *MUC12* expression was observed in all cell lines, although two cell lines, SW620 and SW116 revealed low levels of expression.

These observations, together with the downregulation data, show that although these genes are downregulated in CRCs, they are still detectable in CRC cell lines. In contrast to the normal gastrointestinal tract and IBD tissues, the expression of *MUC11* and *MUC12* in CRCs and in CRC cell lines show patterns of expression distinct form each other.

Referring to FIG. 14, the results of Northern analysis with a

dd34 *(MUC11)* probe showed that in nucleic acid extracts obtained from colonic tissue of four (4) of the (6) CRC patients tested, the level of *MUC11* mRNA expression was lower relative to normal colonic tissue from the same patients. Similarly, *MUC12* mRNA was downregulated in three (3) of five (5) CRC patients (data not shown).

Such quantitative (e.g. downregulation of these genes and differential downregulation expression patterns of *MUC11* and *MUC12)* and also qualitative changes of these genes, e.g. mutations, could be used for diagnostic and prognostic testing in CRC.

EXAMPLE 17: ***Expression of MUC11 and MUC12 in IBD by RT-PCR***

The expression patterns of *MUC11* and *MUC12* in IBD were investigated by RT-PCR, and the results are shown in FIG. 12. Cytokeratin 20, (CK20) a colonic epithelial marker, was employed as a loading control due to the variable epithelial content of IBD tissues. 'N' denotes tissues which appear macroscopically normal and 'D' refers to tissues reported to have IBD. 'CA' refers to the caecum, 'CO' the colon, 'LC' the left colon, 'TC' the transverse colon, 'RS' the recto-sigmoid colon, 'SI' the small intestine, 'IL'denotes the ileum and 'IP' an ileal pouch.

Two patients, patient 1 and patient 4, show 3-4 fold upregulated expression of *MUC11* and *MUC12* in diseased tissues, compared with the same intestinal region observed in the 3 normal controls. Patient 6, who has a history of severe ulcerative colitis in the right colon, also revealed approximately 3-fold upregulated expression of *MUC11* and *MUC12* compared to the right colon observed in the normal controls.

There is coordinate regulation of Mucin expression in the normal gastrointestinal tract as well as in IBD tissues and upregulation of both Mucin genes was observed in 3/10 patients. Given the documented quantitative changes in the expression of *MUC11* and *MUC12*, their expression levels may form the basis of useful diagnostic and prognostic testing for this disease. Qualitative changes in these genes, eg. mutations may also be useful markers for IBD.

EXAMPLE 18: ***Expression of MUC11 and MUC12 in BC by RT-PCR***

The expression patterns of *MUC11* and *MUC12* in BC tissue were investigated by RT-PCR, and the results are shown in FIG. 13. After 40 rounds of amplification, *MUC11* expression was identified in all breast cancer cell lines under investigation; at low levels in BT-20, DU4475, MDA-MB-435 and ZR-75-30 cell lines and at higher levels in the remaining nine cell lines. Eight of the cell lines showed *MUC11* expression higher than the normal colonic cDNA positive control. *MUC11* is clearly highly expressed by most breast cancers and may impact upon the behaviour of the breast cancer cells. *MUC11* may also be secreted by breast cancers and detection in serum could form the basis of diagnostic and prognostic testing for breast cancer. *MUC12* expression was only readily identifiable in one breast cancer cell line, MCF7, although faint bands were observed for BT20, KPL-1 and MA11 cell lines.

EXAMPLE 19: ***Experimental Summary***

Differential display has been used to identify two partial cDNAs, which encode novel colonic mucin-like proteins. Expression of both cDNAs, designated *MUC11* and *MUC12* by the Human Nomenclature Committee, was commonly downregulated in colorectal cancers.

MUC11 and *MUC12* were mapped by FISH to chromosome band 7q22. The location of another mucin gene, *MUC3,* at 7q22, suggests the identification of a new cluster of mucin genes at this locus. Interestingly, four genes encoding gel-forming mucins are found in a cluster on chromosome 11 and these genes appear to have originated from a common ancestral gene. While the mucin cDNAs mapped to 7q22 most likely represent separate genes, it is also possible that they are produced as a result of alternative mRNA splicing from a single, large mucin gene. Northern blot analysis for *MUC11, MUC12* and *MUC3* shows that these encode large transcripts, estimated to be greater than 12 kb.

Multiple tissue RNA analysis showed no cross-reactivity between *MUC11, MUC12* or *MUC3. MUC11* and *MUC12* showed

predominant expression in the colon, while *MUC3* was predominantly expressed in the small intestine and at very low levels in the colon. This expression pattern constitutes an important point of distinction between *MUC11* and *MUC12* genes of the present invention and *MUC3*. Furthermore, the sequences of *MUC11* and *MUC12* are not homologous with any other human mucin genes, but show some degree of similarity within their variable tandem repeat regions to each other (71% over 653 bp). However, their clear differential expression patterns in normal and tumor tissues as well as tumor cell lines, show that they are distinct from each other, and from *MUC3*.

While both MUC11 and MUC12 contain variable repeat regions typical of mucins, MUC12 is putatively a transmembrane mucin with features suggesting an involvement in growth regulation, a largely unrecognized function in human mucins. MUC12 is only the fourth human membrane-anchored epithelial mucin to be described to date, along with MUC1, MUC3 and MUC4. MUC1 has been shown to be involved in cell signaling via multiple tyrosine phosphorylation sites on its highly conserved cytoplasmic tail (Zrihan-Licht *et al.*, 1994, FEBS Lett. **356** 130). At its carboxyl terminus, MUC12 possesses a cytoplasmic tail containing a YNNF sequence (amino acids 557-560 in FIG. 2) which is similar to motifs recognized by SH2 domain-containing proteins (Songyang *et al.*, Mol. Cell. Biol. **14** 2777), suggesting that MUC12, like MUC1, may be involved in signal transduction.

The deduced amino acid sequence of the partial *MUC11* cDNA was composed entirely of serine/threonine-rich tandem repeats. There is a similarity between the tandem repeat consensus sequences of MUC11 (FIG. 4) and MUC12 (FIG. 2) and these also show limited homology to the MUC3 repeat (ITTETTSHSTPSFTSS). These similarities are consistent with evolution from a common ancestral gene. *MUC11* is more widely expressed than *MUC12* and *MUC3* however, with RNA detected in gastrointestinal, respiratory, reproductive and urinary tracts, and unexpectedly in the liver and thymus.

The physiological roles of *MUC11* and *MUC12* in colonic

epithelium are unknown. *MUC11* and *MUC12* are commonly downregulated in colorectal cancer suggesting they may play a role in epithelial cell growth modulation and/or differentiation. At present, it is not possible to comment on whether downregulation of these genes is related to stage of tumor progression, as only 20 patients were analyzed in this study. However, downregulation appears to be so frequent, that it may be an early event in tumorigenesis. Given the co-localization of the *MUC11* and *MUC12* genes on chromosome 7q22, it is possible that their expression is co-ordinately regulated and hence they are simultaneously downregulated in a large proportion of colorectal cancers. The effect of downregulation of these mucins on normal colonic epithelial cells could be substantial. Mucins are believed to protect epithelial cells from attack by pathogenic organisms and from mechanical and chemical damage. Therefore, reduced expression of these mucins could expose colonic epithelial cells to the harsh environment of the intestinal lumen. Furthermore, loss of a transmembrane mucin such as MUC12 may also contribute to loss of critical cell signaling.

The location of these two novel mucin genes on chromosome 7q22 may have significance for two non-malignant epithelial diseases where aberrant mucin expression and/or function is a recognized component of pathology, namely, inflammatory bowel disease and cystic fibrosis. Susceptibility genes for inflammatory bowel disease have been located to chromosomes 3, 12 and 7q22 (Satsangi *et al.*, 1996, Nature Genet. **14** 199). Thus, *MUC11* and *MUC12* must be considered candidates for involvement in inflammatory bowel disease given their chromosomal localization, expression in normal colon, and the documented alterations in mucins in this disease (Rhodes, 1997, QJM **90** 79). Mucins may also play a role in cystic fibrosis as patients with the same *CFTR* gene mutation do not demonstrate exactly the same phenotype in terms of mucus obstruction. The existence of modifier genes has been postulated and mucin genes are obvious candidates (Harris & Reid, 1997, J. Med. Genet. **35** 82). A murine Mucin gene that shows C-terminal homology with *MUC12* has recently been shown

to be a major constituent of obstructive mucus in the gastrointestinal tract of mice with CF (Parmley *et al.*, 1998, J. Clin. Invest **102** 1798).

The *CFTR* gene lies in the adjacent chromosome band (7q31) to the *MUC3, MUC11* and *MUC12* genes. While the significance of these findings is not clear, *MUC11* and *MUC12*, which are expressed in many of the tissues affected by cystic fibrosis, should be considered as candidate modifier genes involved in the aetiology of this disease.

Mucins are encoded by large genes which have proved difficult to clone by conventional methods due to the repetitive nature of their tandem repeat regions. Hereinbefore, the present inventors have unexpectedly identified by differential display two partial cDNAs which represent novel mucin genes that are predominantly expressed in colonic epithelium, both of which are downregulated in colorectal cancer. In this regard, *MUC11* and *MUC12* differ from the other mucin gene located on chromosome 7q22, *MUC3*. These findings together with the sequence homology between the MUC12 EGF-like domain and EGF receptor-binding growth factors, suggest MUC11 and MUC12 may function as growth regulators in colonic epithelium. Downregulation of these two novel mucin genes could be an important and previously unrecognized step in colorectal carcinogenesis.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

TABLES

TABLE 1

NON-CONVENTIONAL AMINO ACID	CODE	NON-CONVENTIONAL AMINO ACID	CODE
α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
aminocyclopropane-	Cpro	L-N-methylasparagine	Nmasn
carboxylate		L-N-methylaspartic acid	Nmasp
aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-	Norb	L-N-methylglutamine	Nmgln
carboxylate		L-N-methylglutamic acid	Nmglu
cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva

NON-CONVENTIONAL AMINO ACID	CODE	NON-CONVENTIONAL AMINO ACID	CODE
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva
D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen

NON-CONVENTIONAL AMINO ACID	CODE	NON-CONVENTIONAL AMINO ACID	CODE
D-α-methylasparagine	Dmasn	α-methyl-α-napthylalanine	Manap
D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
D-α-methylleucine	Dmleu	α-napthylalanine	Anap
D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec

NON-CONVENTIONAL AMINO ACID	CODE	NON-CONVENTIONAL AMINO ACID	CODE
D-α-methylvaline	Dmval	N-cylcododecylglycine	Ncdod
D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser
D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis
D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp
D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro

NON-CONVENTIONAL AMINO ACID	CODE	NON-CONVENTIONAL AMINO ACID	CODE
N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
L-ethylglycine	Etg	penicillamine	Pen
L-homophenylalanine	Hphe	L-α-methylalanine	Mala
L-α-methylarginine	Marg	L-α-methylasparagine	Masn
L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
L-α-methylleucine	Mleu	L-α-methyllysine	Mlys

NON-CONVENTIONAL AMINO ACID	CODE	NON-CONVENTIONAL AMINO ACID	CODE
L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
L-α-methylserine	Mser	L-α-methylthreonine	Mthr
L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr
L-α-methylvaline	Mval	L-N-methylhomophenylalanine	Nmhphe
N-(N-(2,2-diphenylethyl)	Nnbhm	N-(N-(3,3-diphenylpropyl)	Nnbhe
carbamylmethyl)glycine		carbamylmethyl)glycine	
1-carboxy-1-(2,2-diphenyl-	Nmbc	ethylamino)cyclopropane	

TABLE 2

Hybridoma	Immunising peptide	Isotype	Peptide reactivity	Immuno-histochemical reactivity with colon
M11.9	MUC11	IgM	++++	Reacts with paraffin embedded tissue, reactivity enhanced by pre-treatment of sections with neuraminidase which removes sialic acid groups.
M12.15	MUC12	IgM	++++	Reacts with paraffin embedded tissue, reactivity enhanced by pre-treatment of sections with neuraminidase which removes sialic acid groups.

CLAIMS

1. An isolated *MUC* nucleic acid corresponding to a *MUC* gene located on human chromosome 7q22, or a mammalian chromosome structurally or functionally equivalent thereto, which *MUC* gene is normally predominantly expressed in the colon.

2. The isolated *MUC* nucleic acid of Claim 1 which comprises a nucleotide sequence encoding an amino acid sequence which comprises SGLSEESTTSHSSPGSTHTTLSPASTTT(SEQ ID NO :1).

3. The isolated *MUC* nucleic acid of Claim 2, wherein the nucleic acid comprises a nucleotide sequence encoding an amino acid sequence according to SEQ ID NO:3.

4. The isolated *MUC* nucleic acid of Claim 2, wherein the nucleic acid includes a sequence of nucleotides according to SEQ ID NO: 2.

5. The isolated *MUC* nucleic acid of Claim 1, wherein the nucleic acid which comprises a nucleotide sequence which encodes an amino acid sequence whuch comprises SGLSQESTTFHSSPGSTETTLAPASTTT (SEQ ID NO: 4).

6. The isolated *MUC* nucleic acid of Claim 5, wherein the nucleic acid comprises a nucleotide sequence which encodes an amino acid sequence according to SEQ ID NO:6.

7. The isolated *MUC* nucleic acid of Claim 5, wherein the nucleic acid includes a sequence of nucleotides according to SEQ ID NO: 5.

8. A *MUC* nucleic acid homolog which hybridizes to the isolated *MUC* nucleic acid of any one of Claims 2-7 under conditions of at least low stringency.

9. A *MUC* nucleic acid homolog which has at least 60% nucleotide sequence identity with the isolated *MUC* nucleic acid of any one of Claims 2-7.

10. An isolated MUC polypeptide having an amino acid sequence according to SEQ ID NO: 3.

11. An isolated MUC polypeptide having an amino acid sequence

according to SEQ ID NO: 6.

12. An isolated MUC polypeptide homolog which has at least 60% amino acid identity with the MUC polypeptide of Claim 10 or Claim 11.

13. An antibody specific for the MUC polypeptide or MUC polypeptide homolog of any one of Claims 10-12.

14. An antibody according to Claim 13 which is a monoclonal antibody.

15. A monoclonal antibody according to Claim 14, which monoclonal antibody is selected from the group consisting of M11.9 and M12.15.

16. A method of detecting the MUC polypeptide of Claim 10 or Claim 11, including the steps of:-

(i) obtaining a sample from said mammal;

(ii) forming a complex between said MUC polypeptide, if present in said sample, and an anti-MUC polypeptide antibody; and

(iii) detecting said MUC polypeptide in said complex.

17. The method of Claim 17, wherein the antibody is selected from the group consisting of M11.9 and M12.15.

18. A method of detecting a *MUC* gene or a *MUC* gene transcript including the steps of:-

(i) obtaining a nucleic acid extract from said mammal;

(ii) forming a hybrid nucleic acid comprising a *MUC* gene or a *MUC* gene transcript if present in said sample, and a corresponding isolated *MUC* nucleic acid according to Claim 1, or a portion thereof; and

(iii) detecting said hybrid nucleic acid.

19. The method of Claim 18, wherein the isolated *MUC* nucleic acid has a nucleotide sequence selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:5.

20. A method of detecting a *MUC* gene or a *MUC* gene transcript

including the steps of:-

(i) obtaining a nucleic acid extract from said mammal;

(ii) using one or more primers, each having a nucleotide sequence corresponding to a distinct portion of the isolated *MUC* nucleic acid of Claim 1, together with a polynucleotide sequence amplification technique, to produce a *MUC* gene amplification product from said extract; and

(iii) detecting said *MUC* gene amplification product.

21. The method of Claim 20, wherein the one or more primers is selected from the group consisting of SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 14 and SEQ ID NO:15.

22. The method of Claim 21, wherein the polynucleotide sequence amplification technique is RT-PCR.

23. Use of the isolated *MUC* nucleic acid of Claim 1, or a portion thereof, to detect a polymorphism, mutation, deletion, truncation and/or expansion in a corresponding *MUC* gene or *MUC* gene transcript.

24. A pharmaceutical composition comprising a pharmaceutically acceptable amount of the MUC polypeptide of Claim 10 or Claim 11, together with a pharmaceutically-acceptable carrier and/or diluent.

25. A pharmaceutical composition comprising a pharmaceutically acceptable amount of an anti-MUC antibody according to any one of Claims 13-15, together with a pharmaceutically-acceptable carrier and/or diluent.

26. A method of treating of a mammal suffering from a disease condition, said method including the step of administering to said mammal a pharmaceutical composition according to Claim 24 or Claim 25 to thereby alleviate or prevent one or more symptoms of said disease condition in said mammal.

27. A method of gene therapy of a mammal suffering from a disease condition, said method including the step of administering a gene therapy construct to said mammal, said gene therapy construct comprising

the isolated *MUC* nucleic acid of Claim 1, or a portion thereof, to thereby alleviate or prevent one or more symptoms of said disease condition in said mammal.

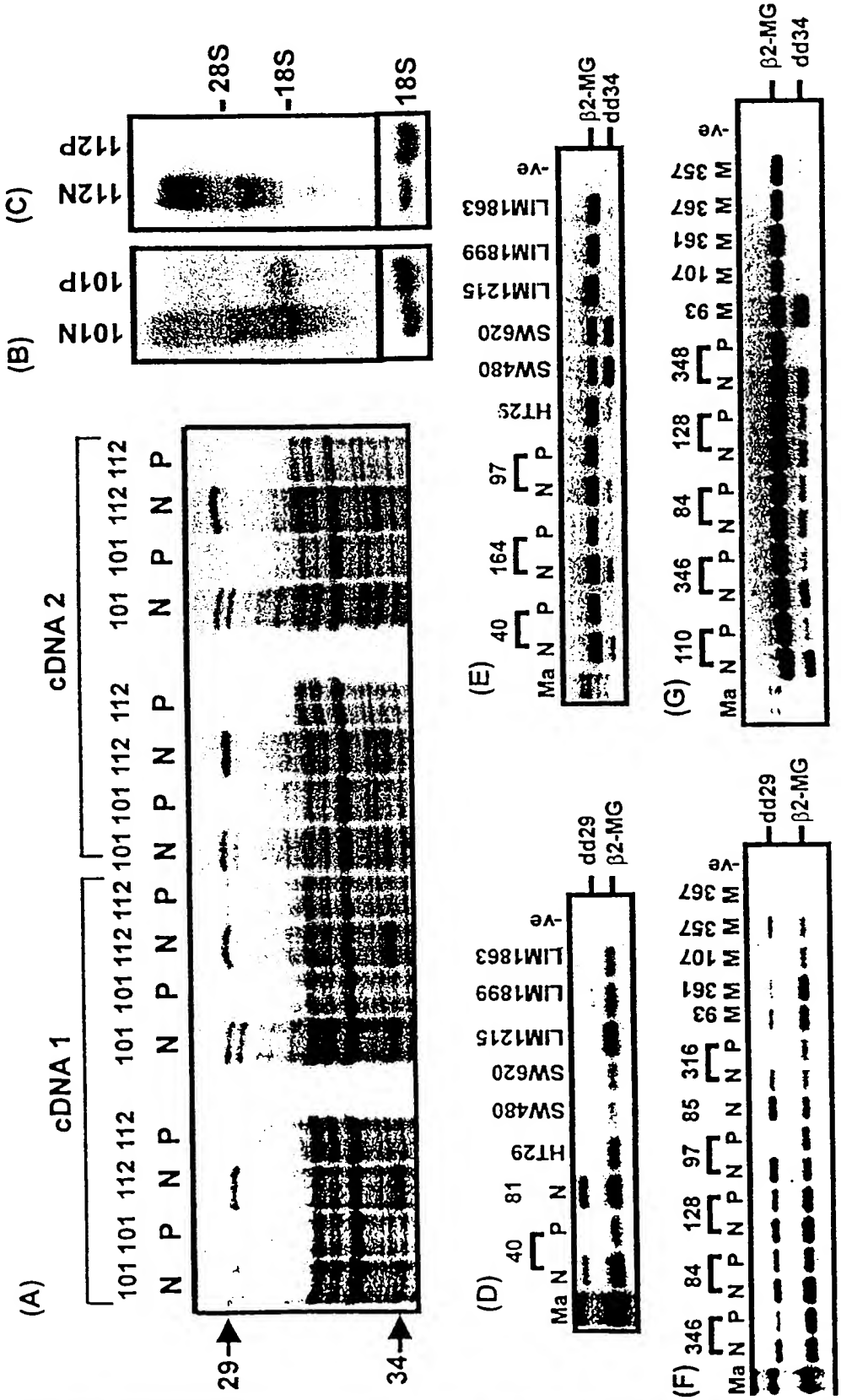
28. The method of Claim 27, wherein the isolated *MUC* nucleic acid has a nucleotide sequence selected from the group consisting of SEQ ID NO:2 and SEQ ID NO:5.

29. The method of any one of Claims 26-28, wherein said disease condition is associated with aberrant Mucin expression, altered properties of mucus or epithelial inflammatory processes involving Mucins

30. The method of any one of Claims 26-28, wherein said disease condition is selected from the group consisting of colorectal cancer (CRC), cystic fibrosis (CF), inflammatory bowel disease (IBD), breast cancer (BC), Crohn's disease, ulcerative colitis, asthma and chronic bronchitis.

31. The method of Claim 30, wherein said disease condition is selected from the group consisting of colorectal cancer (CRC), cystic fibrosis (CF), inflammatory bowel disease (IBD) and breast cancer (BC).

32. The method of any one of Claims 26-31, wherein the mammal is a human.


(A)
cDNA 1
cDNA 2
N P N P N P N P N P N P N P N P
101 101 112 112 101 101 112 112 101 101 112 112 101 101 112 112
29
34
(B)
101N
101P
(C)
112N
112P
28S
18S
18S
(D)
Ma N P N
40
81
HT29
SW480
SW620
LIM1215
LIM1899
LIM1863
-ve
dd29
β2-MG
(E)
Ma N P N P N P
40
164
97
HT29
SW480
SW620
LIM1215
LIM1899
LIM1863
-ve
β2-MG
dd34
(F)
Ma N P N P N P N P N P
346
84
128
97
85
316
93
361
107
357
367
-ve
dd29
β2-MG
(G)
Ma N P N P N P N P N P
110
346
84
128
348
93
361
107
357
367
-ve
β2-MG
dd34

Consensus tandem repeat sequence

S G L S Q E S T T F H S S P G S T E T T L S P A S T T T
H S

```
                                    T L S P A S M R S   9
S S I S G E P T S L Y S Q A E S T H T T A F P A S T T T  37
S G L S Q E S T T F H S K P G S T E T T L S P G S I T T  65
S S F A Q E F T T P H S Q P G S A L S T V S P A S T T V  93
P G L S E E S T T F Y S S P G S T E T T A F S H S N T M 121
S I H S Q Q S T P F P D S P G F T H T V L - P A T L T T 148
T D I G Q E S T A F H S S S D A T G T T P L P A R S T A 176
S D L V G E P T T F Y I S P S P T Y T T L F P A S S S T 204
S G L T E E S T T F H T S P S F T S T I V S T E S L E T 232
L A P G L C Q E G Q I W N G K Q C V C P Q G Y V G Y Q C 260
L S P L E S F P V E T P E K L N A T L G M T V K V T Y R 288
N F T E K M N D A S S Q E Y Q N F S T L F K N R M D V V 316
L K G D N L P Q Y R G V N I R R L L N G S I V V K N D V 344
I L E A D Y T L E Y E E L F E N L A E I V K A K I M N E 372
T R T T L L D P D S C R K A I L C Y S E E D T F V D S S 400
V T P G F D F Q E Q C T Q K A A E G Y T Q F Y Y V D V L 428
D G K L A C V N K C T K G T K S Q M N C N L G T C Q L Q 456
R S G P R C L C P N T N T H W Y W G E T C E F N I A K S 484
L V Y G I V G A V M A V L L L A L I I L I I L F S L S Q 512
R K R H R E Q Y D V P Q E W R K E G T P G I F Q K T A I 540
W E D Q N L R E S R F G L E N A Y N N F R P T L E T V D 568
S G T E L H I Q R P E M V A S T V *                   585
```

FIG. 2

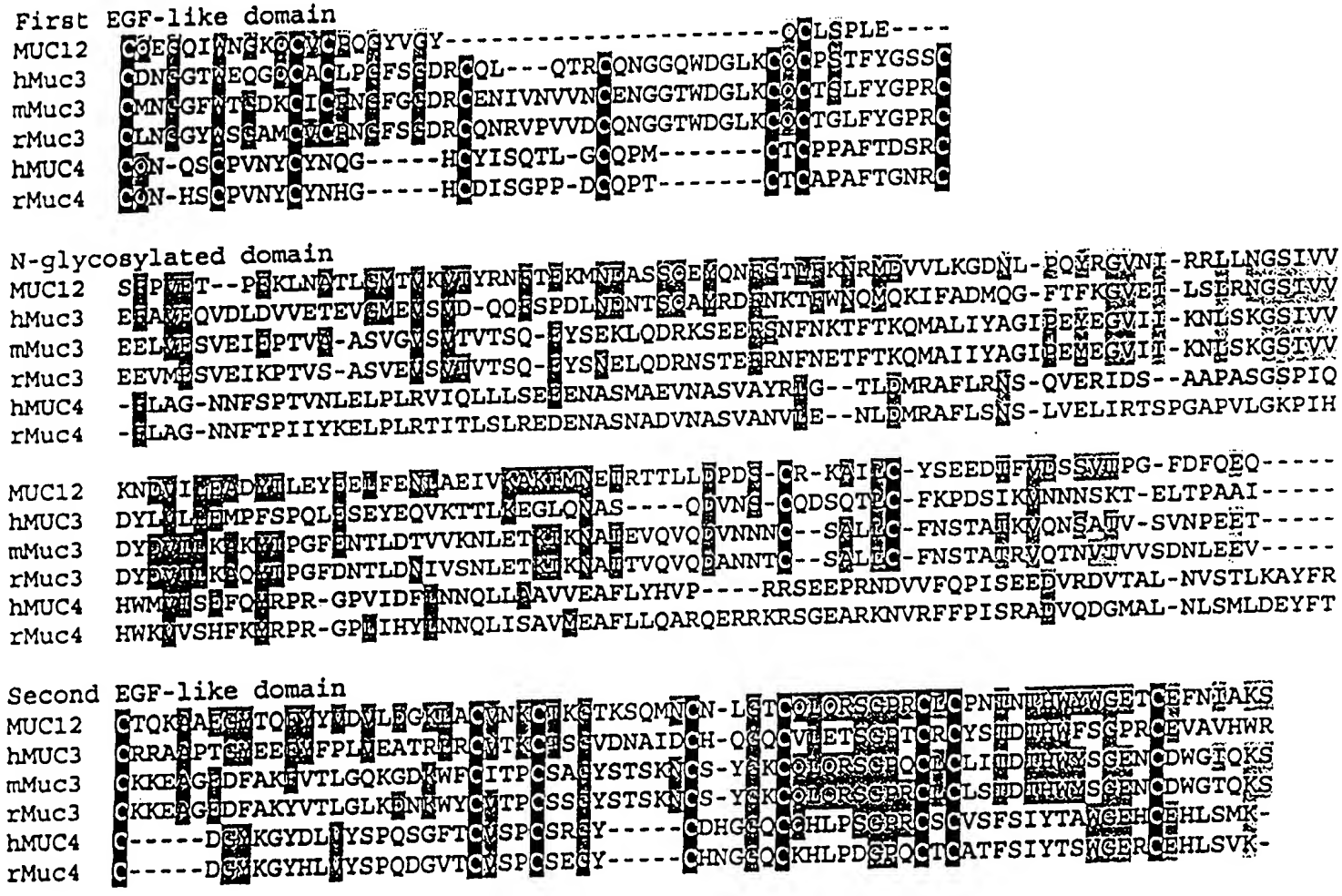
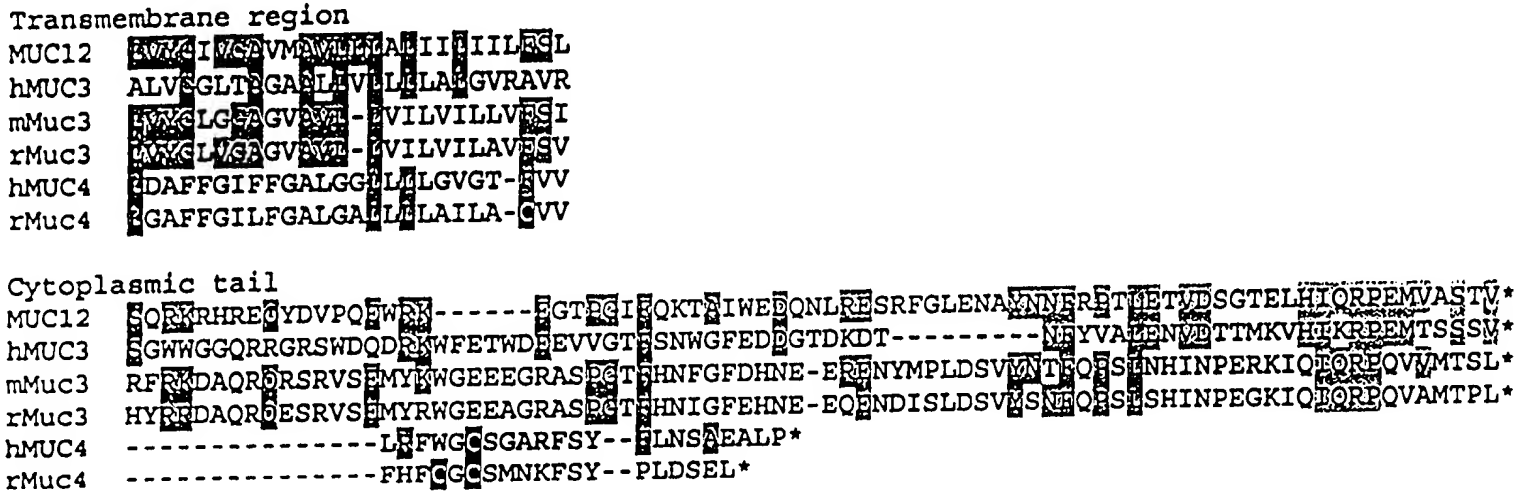

First EGF-like domain
MUC12
hMuc3
mMuc3
rMuc3
hMUC4
rMuc4
N-glycosylated domain
MUC12
hMuc3
mMuc3
rMuc3
hMUC4
rMuc4
MUC12
hMUC3
mMuc3
rMuc3
hMUC4
rMuc4
Second EGF-like domain
MUC12
hMUC3
mMuc3
rMuc3
hMUC4
rMuc4

FIG. 3

FIG. 3 continued



Consensus tandem repeat sequence
SGLSEESTTSHSSPGSTHTTLSPASTTT
F

RNRPHTTAFPGSTTM 15
PGVSQESTASHSSPGSTDTTLSPGSTTA 43
SSLGPESTTFHSGPGSTETTLLPDNTTA 71
SGLLEASTPVHSSTGSPHTTLSPAGSTT 99
--RQGESTTFQSWPNSKDTTPAP-PTTT 124
SAFVELSTTSHGSPSSTPTTHFSASSTT 152
LGRSEESTTVHSSPVATATTPSPARSTT 180
SGLVEESTTYHSSPGSTQTMHFPESDTT 208
SGRGEESTTSHSS--TTHTI-SSAPSTT 233
SALVEEPTSYHSSPGSTATTHFPDSSTT 261
SGRSEESTASHSNQDATGTIVLPARSTT 289
SVLLGESTTSPISSGSMETTALPGSTTT 317
PGLSEKSTTFHSSPRSPATTLSPASTTS 345
SGVSEESTTSHSRPGSTHTTAFPDSTTT 373
PGLSRHSTTSHSSPGSTDTTLLPASTTT 401
SGPSQESTTSHSSPGSTDTALSPGSTTA 429
LSFGQESTTFHSSPGSTHTTLFPDSTTS 457
SGIVEASTRVHSSTGSPRTTLSPASSTS 485
PGLQGESTAFQTHPASTHTTPSTPSTAT 513
-APVEESTTYHRSPSSTPTTHFPASSTT 540
SGHSEKSTIFHSSPDASGTTPSSAHSTT 568
SGRG-ESTTSRISPGSTEITTLPGSTTT 595
PGLSEASTTFYSSPRSPTTTLSPASMTS 623
LGVGEESTTSRSQPGSTHSTVSPASTTT 651
PGLSEESTTVYSSSPGSTETTVFPRSTT 679
TSVRGEEPTTFHSRPASTHTTLFTEDST 707
TSGLTEESTAFPGSPASTQTGLPATLTT 735
ADLGEESTTFPSSSGSTGTTLSPARSTT 763
SGLVGESTPSRLSPSSTETTTLPGSPTT 791
PSLSEKSTTFYTSPRSPDATLSPATTTS 819
SGVSEESSTSHSQPGSTHTTAFPDSTTT 847
SGLSQEPKTSHSSQGSTEATLSPGSTTA 875
SSLGQQSTTFHSSPGDTETTLLPDDTIT 903
SGLVEASTPTHSSTGSLHTTLTPASSTS 931
AGLQEESTTFQSWPSSSDTTPSPPGP 957

FIG. 4

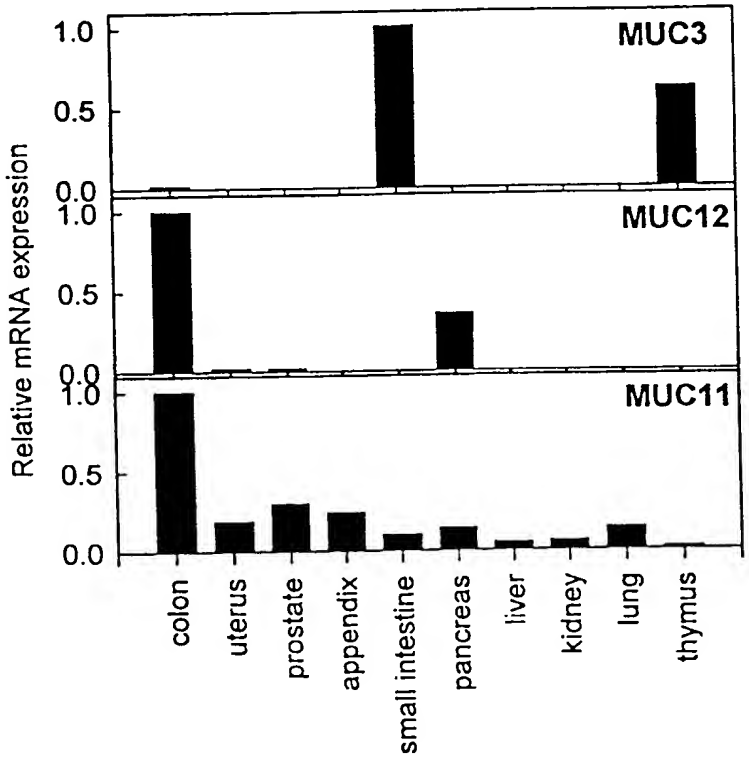

MUC3
MUC12
MUC11
1.0
0.5
0.0
Relative mRNA expression
colon
uterus
prostate
appendix
small intestine
pancreas
liver
kidney
lung
thymus

FIG. 5

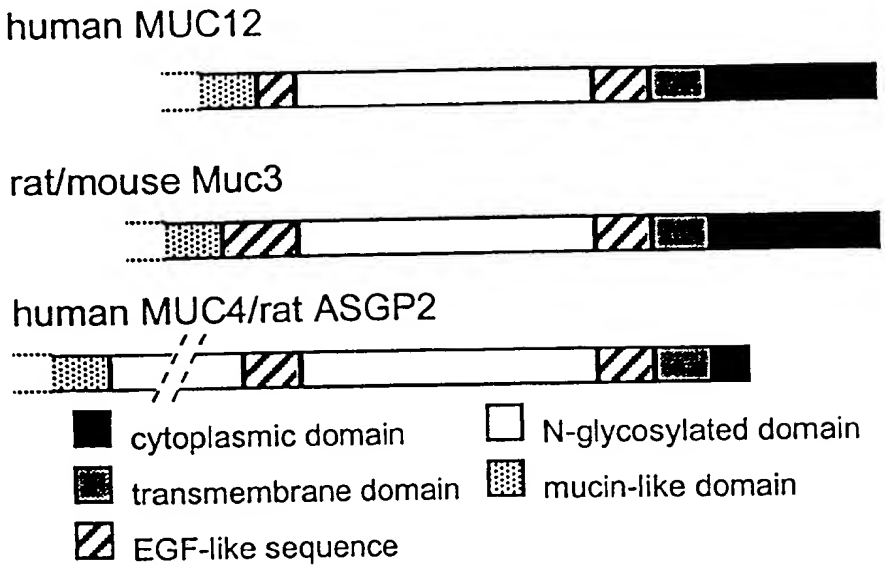

human MUC12
rat/mouse Muc3
human MUC4/rat ASGP2
cytoplasmic domain
N-glycosylated domain
transmembrane domain
mucin-like domain
EGF-like sequence

FIG. 6

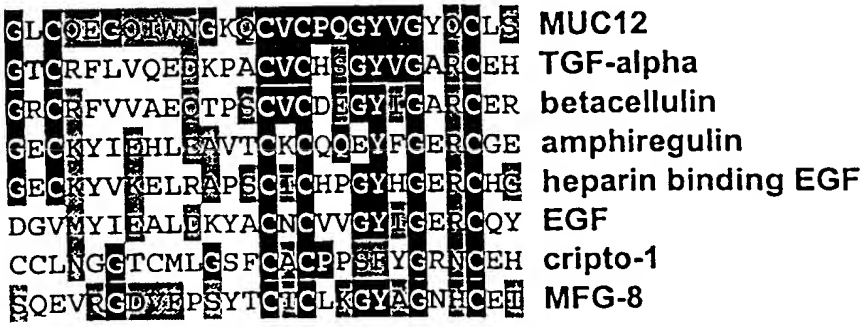

MUC12
TGF-alpha
betacellulin
amphiregulin
heparin binding EGF
EGF
cripto-1
MFG-8

FIG. 7

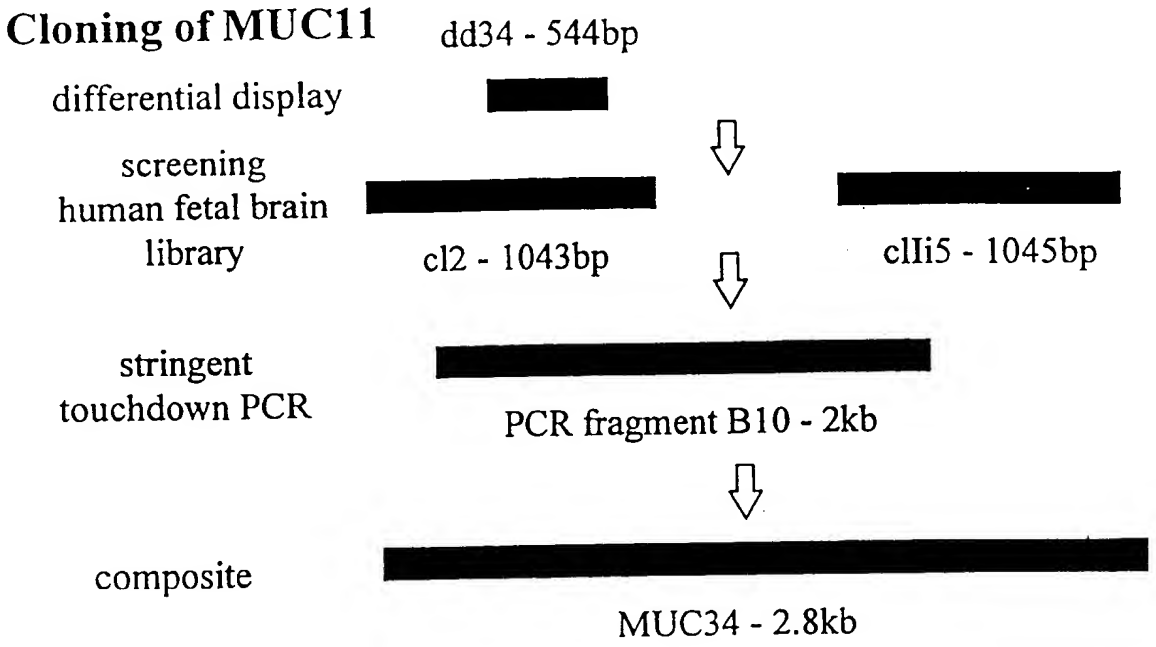

Cloning of MUC11
dd34 - 544bp
differential display
screening
human fetal brain
library
cl2 - 1043bp
clIi5 - 1045bp
stringent
touchdown PCR
PCR fragment B10 - 2kb
composite
MUC34 - 2.8kb

FIG. 8A

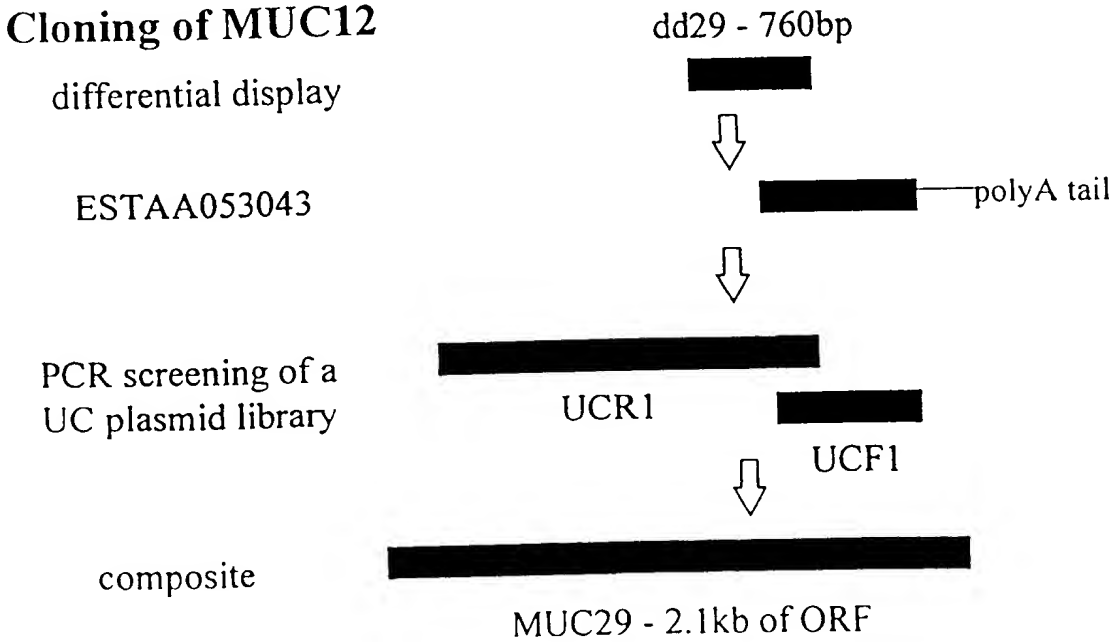
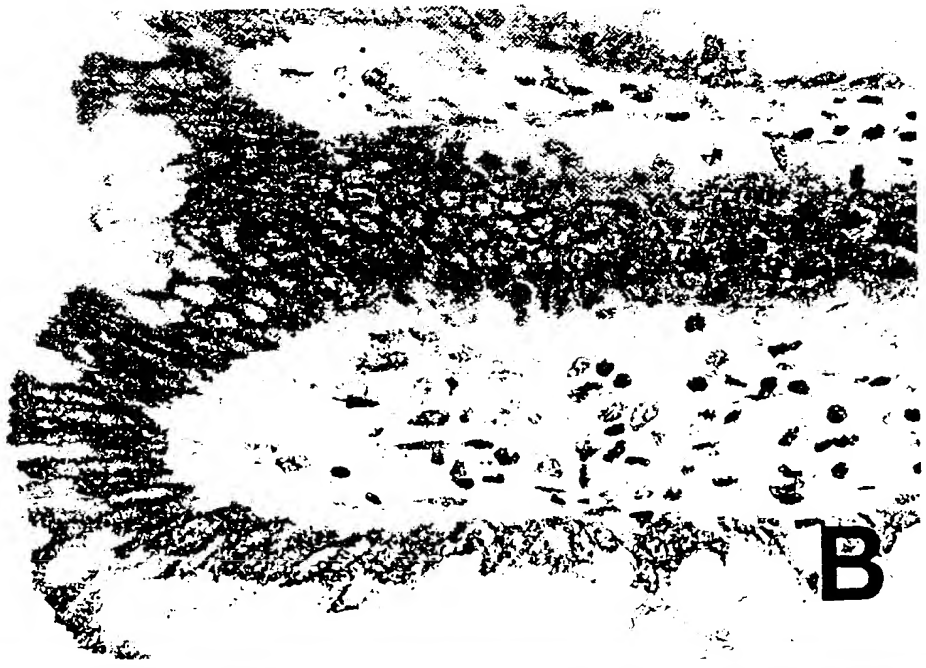

Cloning of MUC12
dd29 - 760bp
differential display
ESTAA053043
polyA tail
PCR screening of a
UC plasmid library
UCR1
UCF1
composite
MUC29 - 2.1kb of ORF

FIG. 8B

FIG. 9A



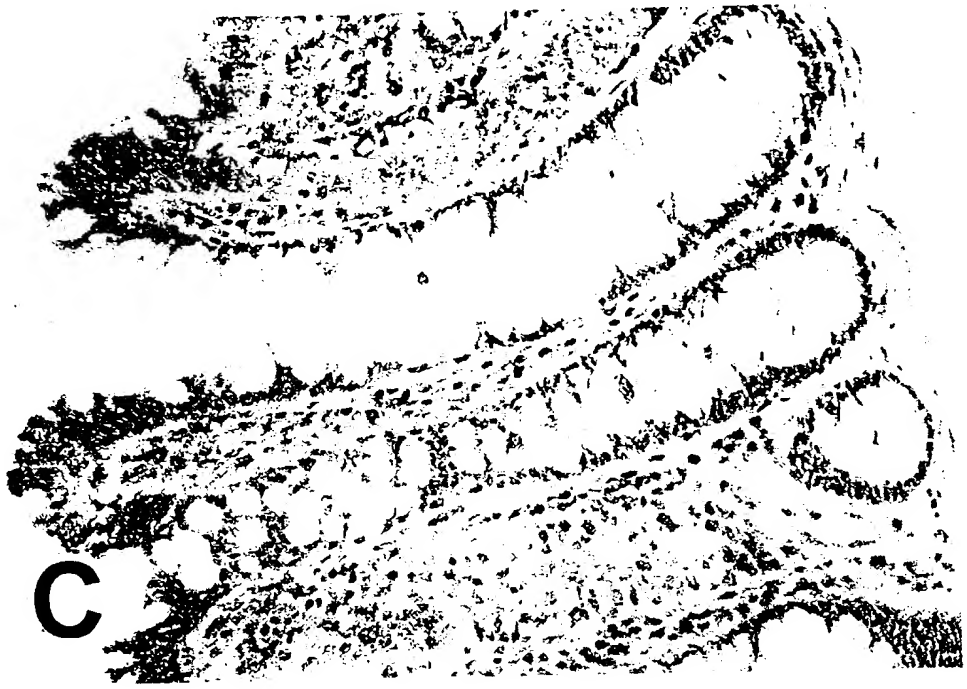


FIG. 9C

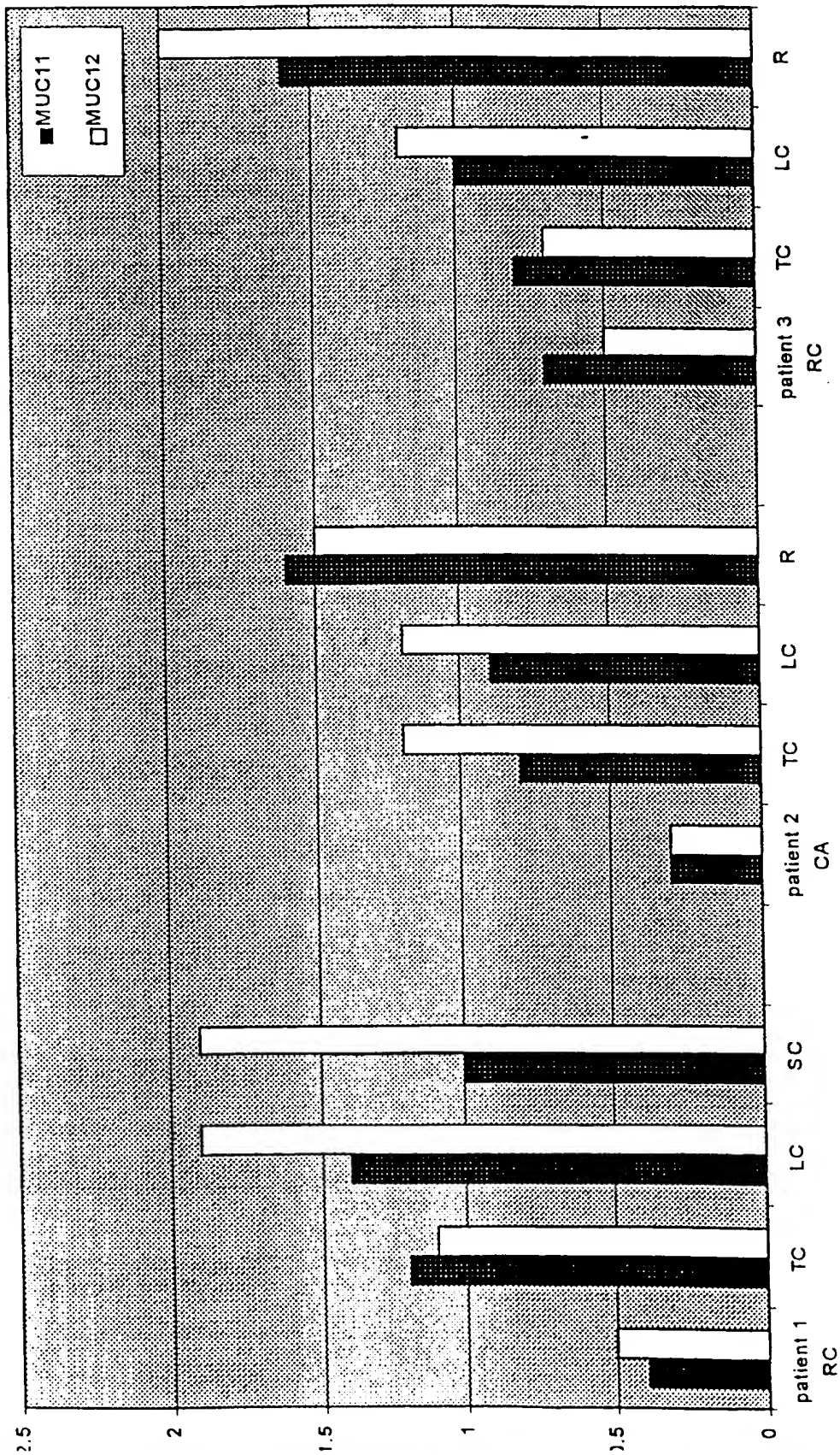

MUC11
MUC12
2.5
2
1.5
1
0.5
0
patient 1
RC
TC
LC
SC
patient 2
CA
TC
LC
R
patient 3
RC
TC
LC
R
Colonic tissues

FIG. 10

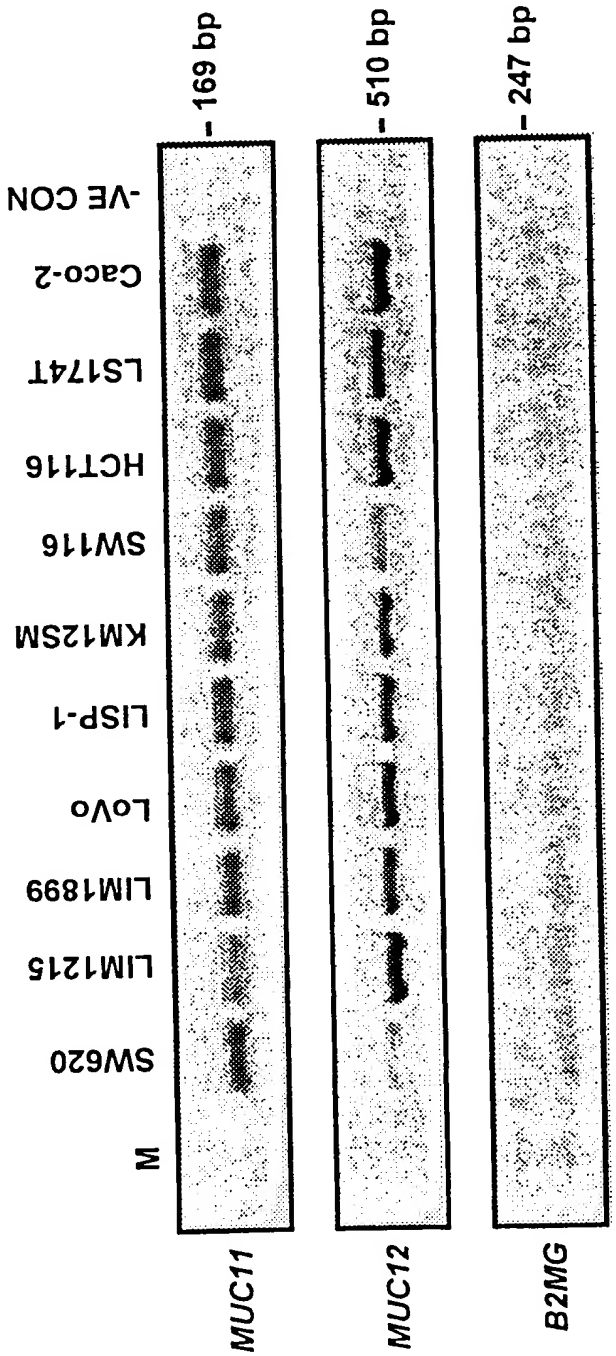

169 bp
510 bp
247 bp
-VE CON
Caco-2
LS174T
HCT116
SW116
KM12SM
LISP-1
LoVo
LIM1899
LIM1215
SW620
M
MUC11
MUC12
B2MG

FIG. 11

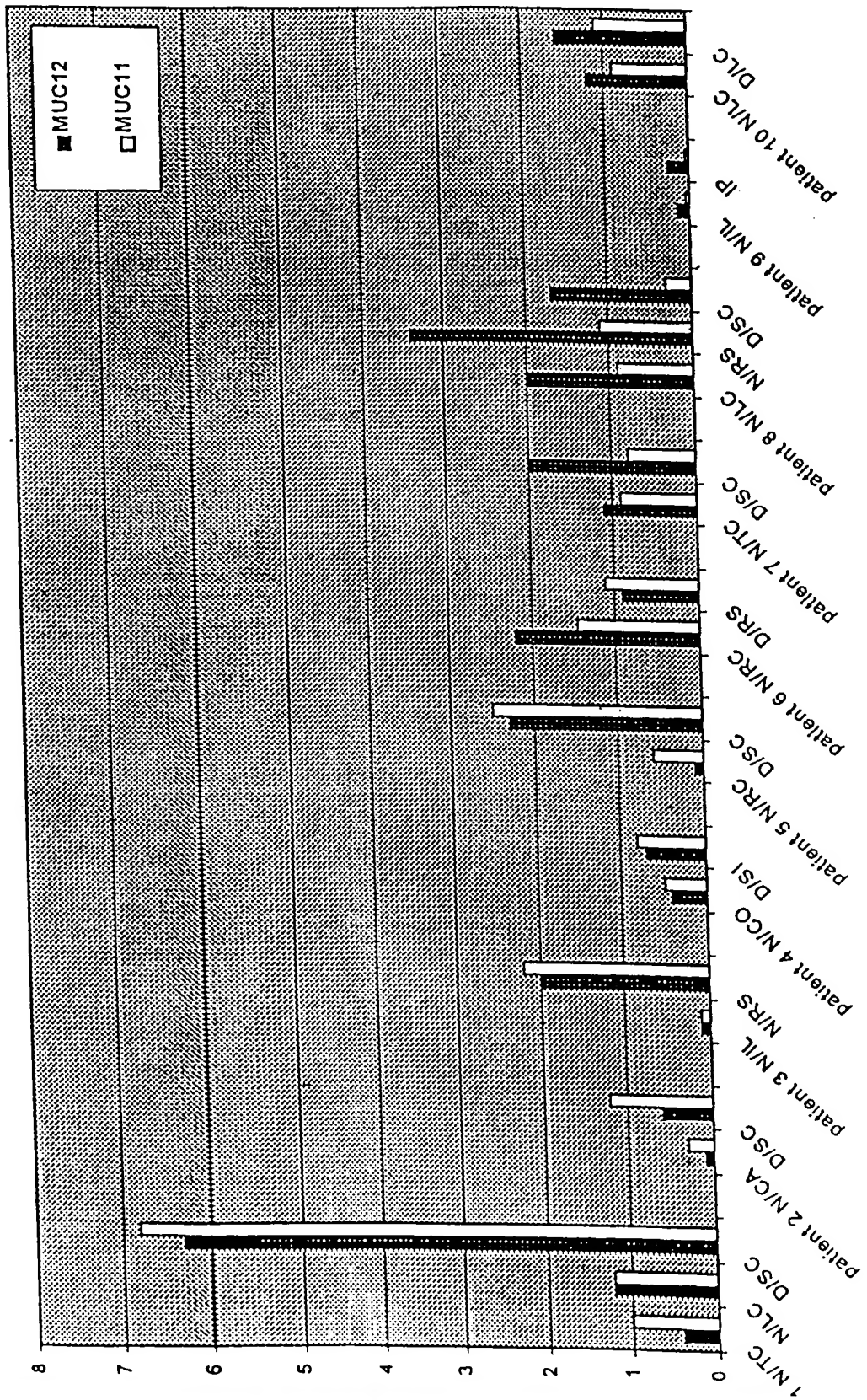

MUC12
MUC11
0
1
2
3
4
5
6
7
8
1 N/TC
N/LC
D/SC
patient 2 N/CA
D/SC
patient 3 N/IL
N/RS
patient 4 N/CO
D/SI
patient 5 N/RC
D/SC
patient 6 N/RC
D/RS
patient 7 N/TC
D/SC
patient 8 N/LC
N/RS
D/SC
patient 9 N/IL
IP
patient 10 N/LC
D/LC
Tissues

FIG. 12

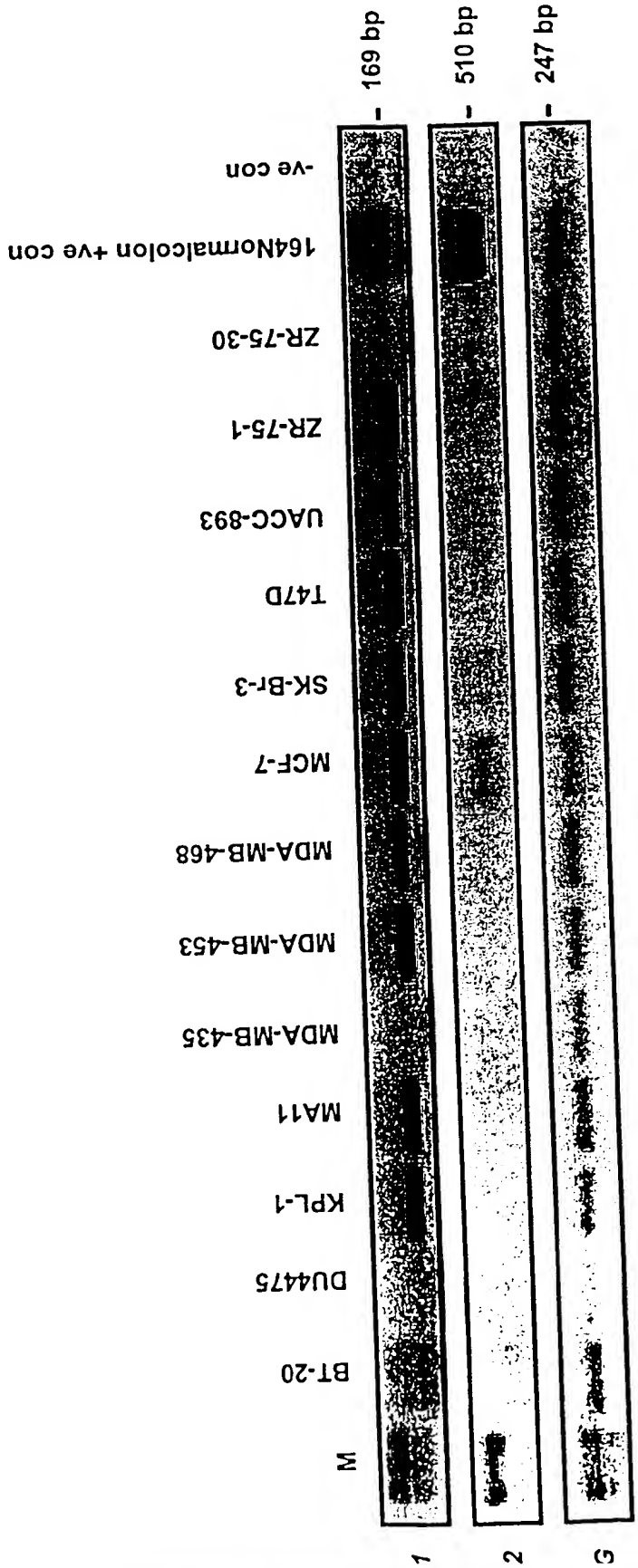

169 bp
510 bp
247 bp
-ve con
164Normalcolon +ve con
ZR-75-30
ZR-75-1
UACC-893
T47D
SK-Br-3
MCF-7
MDA-MB-468
MDA-MB-453
MDA-MB-435
MA11
KPL-1
DU4475
BT-20
M
1
2
3

FIG. 13

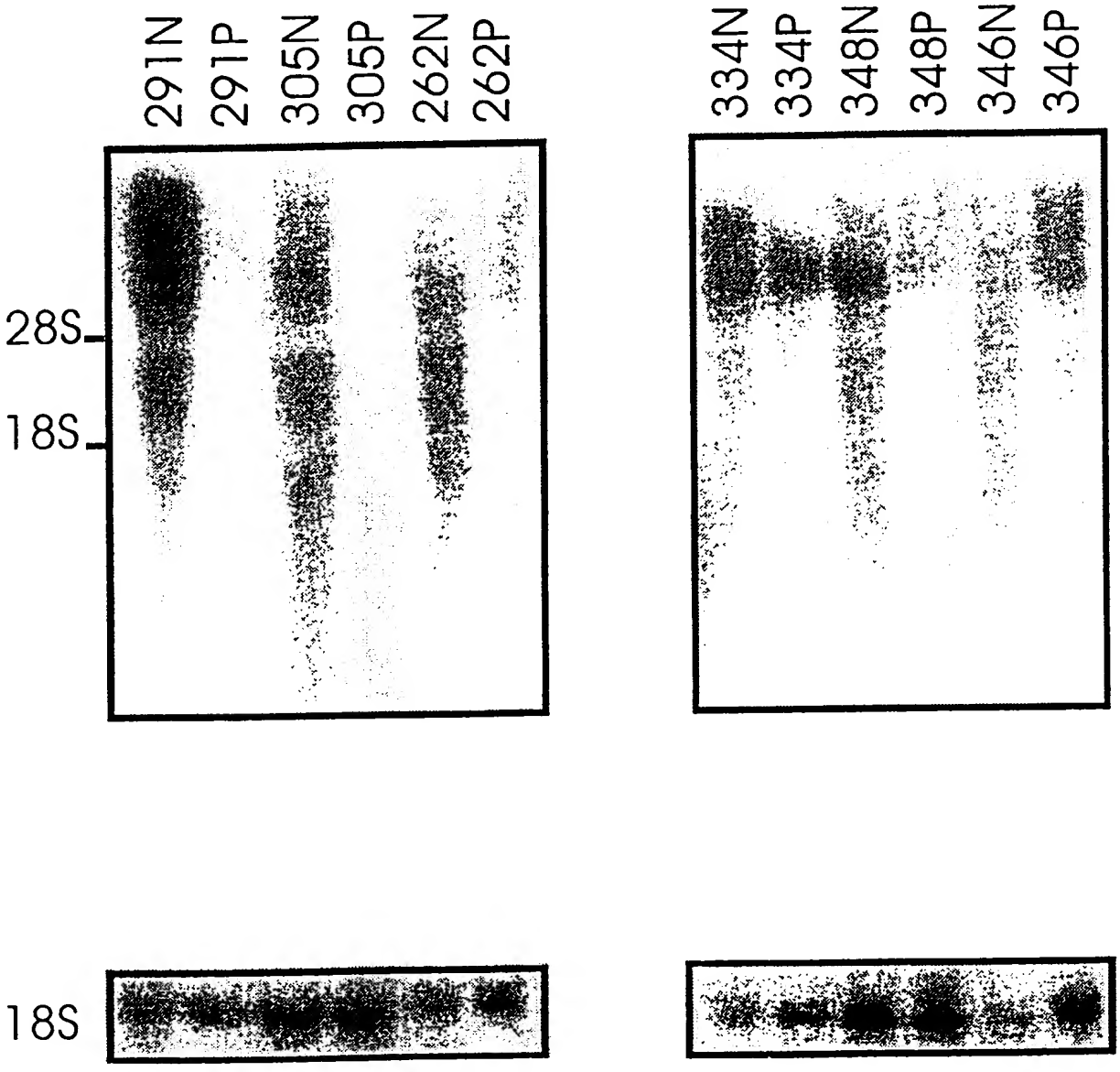

291N
291P
305N
305P
262N
262P
334N
334P
348N
348P
346N
346P
28S
18S
18S

FIG. 14

SEQUENCE LISTING

<110> The Council of the Queensland Institute of Medical
The Corporation of the Trustees of the Order of th

<120> MUCINS

<130> mucinqimr

<140>
<141>

<150> PP4708
<151> 1998-07-16

<160> 18

<170> PatentIn Ver. 2.0

<210> 1
<211> 28
<212> PRT
<213> Homo sapiens

<220>
<221> REPEAT
<222> (1)..(28)
<223> MUC11 consensus tandem repeat sequence

<400> 1
Ser Gly Leu Ser Glu Glu Ser Thr Thr Ser His Ser Ser Pro Gly Ser
1 5 10 15

Thr His Thr Thr Leu Ser Pro Ala Ser Thr Thr Thr
20 25

<210> 2
<211> 2872
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(2871)

<400> 2

```
Arg Asn Arg Pro His Thr Thr Ala Phe Pro Gly Ser Thr Thr Met Pro
 1               5                  10                  15

ggc gtc agt cag gaa tct aca gct tcc cac agc agc cca ggc tcc aca   96
Gly Val Ser Gln Glu Ser Thr Ala Ser His Ser Ser Pro Gly Ser Thr
            20                  25                  30

gac aca aca ctg tcc cct ggc agt acc aca gca tca tcc ctt ggt cca   144
Asp Thr Thr Leu Ser Pro Gly Ser Thr Thr Ala Ser Ser Leu Gly Pro
        35                  40                  45

gaa tct act acc ttc cac agc ggc cca ggc tcc act gaa aca aca ctc   192
Glu Ser Thr Thr Phe His Ser Gly Pro Gly Ser Thr Glu Thr Thr Leu
    50                  55                  60

tta cct gac aac acc aca gcc tcc ggc ctc ctt gaa gca tct acg ccc   240
Leu Pro Asp Asn Thr Thr Ala Ser Gly Leu Leu Glu Ala Ser Thr Pro
65                  70                  75                  80

gtc cac agc agc act gga tcg cca cac aca aca ctg tcc cct gcc ggc   288
Val His Ser Ser Thr Gly Ser Pro His Thr Thr Leu Ser Pro Ala Gly
                85                  90                  95

tct aca acc cgt cag gga gaa tct acc acc ttc cag agc tgg cct aac   336
Ser Thr Thr Arg Gln Gly Glu Ser Thr Thr Phe Gln Ser Trp Pro Asn
            100                 105                 110

tcg aag gac act acc cct gca cct cct act acc aca tca gcc ttt gtt   384
Ser Lys Asp Thr Thr Pro Ala Pro Pro Thr Thr Thr Ser Ala Phe Val
        115                 120                 125

gag cta tct aca acc tcc cac ggc agc ccg agc tca act cca aca acc   432
Glu Leu Ser Thr Thr Ser His Gly Ser Pro Ser Ser Thr Pro Thr Thr
    130                 135                 140

cac ttt tct gcc agc tcc aca acc ttg ggc cgt agt gag gaa tcg aca   480
His Phe Ser Ala Ser Ser Thr Thr Leu Gly Arg Ser Glu Glu Ser Thr
145                 150                 155                 160

aca gtc cac agc agc cca gtt gca act gca aca aca ccc tcg cct gcc   528
Thr Val His Ser Ser Pro Val Ala Thr Ala Thr Thr Pro Ser Pro Ala
                165                 170                 175

cgc tcc aca acc tca ggc ctc gtt gaa gaa tct acg acc tac cac agc   576
Arg Ser Thr Thr Ser Gly Leu Val Glu Glu Ser Thr Thr Tyr His Ser
            180                 185                 190

                                                                  624
```

```
Ser Pro Gly Ser Thr Gln Thr Met His Phe Pro Glu Ser Asp Thr Thr
        195                 200                 205

tca ggc cgt ggt gaa gaa tca aca act tcc cac agc agc aca aca cac   672
Ser Gly Arg Gly Glu Glu Ser Thr Thr Ser His Ser Ser Thr Thr His
    210                 215                 220

aca ata tct tca gct cct agc acc aca tct gcc ctt gtt gaa gaa cct   720
Thr Ile Ser Ser Ala Pro Ser Thr Thr Ser Ala Leu Val Glu Glu Pro
225                 230                 235                 240

acc agc tac cac agc agc ccg ggc tca act gca aca aca cac ttc cct   768
Thr Ser Tyr His Ser Ser Pro Gly Ser Thr Ala Thr Thr His Phe Pro
                245                 250                 255

gac agc tcc aca acc tca ggc cgt agt gag gaa tca aca gca tcc cac   816
Asp Ser Ser Thr Thr Ser Gly Arg Ser Glu Glu Ser Thr Ala Ser His
            260                 265                 270

agc aac caa gac gca acg gga aca ata gtc cta cct gcc cgc tcc aca   864
Ser Asn Gln Asp Ala Thr Gly Thr Ile Val Leu Pro Ala Arg Ser Thr
        275                 280                 285

acc tca gtt ctt ctt gga gaa tct acg acc tca ccc atc agt tca ggc   912
Thr Ser Val Leu Leu Gly Glu Ser Thr Thr Ser Pro Ile Ser Ser Gly
    290                 295                 300

tca atg gaa acg aca gcg tta ccc ggc agt acc aca acg cca ggc ctc   960
Ser Met Glu Thr Thr Ala Leu Pro Gly Ser Thr Thr Thr Pro Gly Leu
305                 310                 315                 320

agt gag aaa tct acc act ttc cac agt agc ccc aga tca cca gcc aca  1008
Ser Glu Lys Ser Thr Thr Phe His Ser Ser Pro Arg Ser Pro Ala Thr
                325                 330                 335

aca ctc tca cct gcc agc acg aca agc tca ggc gtc agt gaa gaa tcc  1056
Thr Leu Ser Pro Ala Ser Thr Thr Ser Ser Gly Val Ser Glu Glu Ser
            340                 345                 350

acc acc tcc cac agc cga cca ggc tca acg cac aca aca gca ttc cct  1104
Thr Thr Ser His Ser Arg Pro Gly Ser Thr His Thr Thr Ala Phe Pro
        355                 360                 365

gac agc acc acc acg cca ggc ctc agt cgg cat tct aca act tcc cac  1152
Asp Ser Thr Thr Thr Pro Gly Leu Ser Arg His Ser Thr Thr Ser His
    370                 375                 380

                                                                 1200
```

```
Ser Ser Pro Gly Ser Thr Asp Thr Thr Leu Leu Pro Ala Ser Thr Thr
385                 390                 395                 400

acc tca ggc ccc agt cag gaa tca aca act tcc cac agc agc cca ggt   1248
Thr Ser Gly Pro Ser Gln Glu Ser Thr Thr Ser His Ser Ser Pro Gly
                405                 410                 415

tca act gac aca gca ctg tcc cct ggc agt acc aca gcc tta tcc ttt   1296
Ser Thr Asp Thr Ala Leu Ser Pro Gly Ser Thr Thr Ala Leu Ser Phe
            420                 425                 430

ggt caa gaa tct aca acc ttc cac agc agc cca ggc tcc act cac aca   1344
Gly Gln Glu Ser Thr Thr Phe His Ser Ser Pro Gly Ser Thr His Thr
        435                 440                 445

aca ctc ttc cct gac agc acc aca agc tca ggc atc gtt gaa gca tct   1392
Thr Leu Phe Pro Asp Ser Thr Thr Ser Ser Gly Ile Val Glu Ala Ser
    450                 455                 460

aca cgc gtc cac agc agc act ggc tca cca cgc aca aca ctg tcc cct   1440
Thr Arg Val His Ser Ser Thr Gly Ser Pro Arg Thr Thr Leu Ser Pro
465                 470                 475                 480

gcc agc tcc aca agc cct gga ctt cag gga gaa tct acc gcc ttc cag   1488
Ala Ser Ser Thr Ser Pro Gly Leu Gln Gly Glu Ser Thr Ala Phe Gln
                485                 490                 495

acc cac cca gcc tca act cac acg acg cct tca act cct agc acc gca   1536
Thr His Pro Ala Ser Thr His Thr Thr Pro Ser Thr Pro Ser Thr Ala
            500                 505                 510

aca gcc cct gtt gaa gaa tct aca acc tac cac cgc agc cca agc tcg   1584
Thr Ala Pro Val Glu Glu Ser Thr Thr Tyr His Arg Ser Pro Ser Ser
        515                 520                 525

act cca aca aca cac ttc cct gcc agc tcc aca act tcg ggc cac agt   1632
Thr Pro Thr Thr His Phe Pro Ala Ser Ser Thr Thr Ser Gly His Ser
    530                 535                 540

gag aaa tca aca ata ttc cac agc agc cca gat gca agt gga aca aca   1680
Glu Lys Ser Thr Ile Phe His Ser Ser Pro Asp Ala Ser Gly Thr Thr
545                 550                 555                 560

ccc tca tct gcc cac tcc aca acc tca ggt cgt gga gaa tct aca acc   1728
Pro Ser Ser Ala His Ser Thr Thr Ser Gly Arg Gly Glu Ser Thr Thr
                565                 570                 575

                                                                  1776
```

```
Ser Arg Ile Ser Pro Gly Ser Thr Glu Ile Thr Thr Leu Pro Gly Ser
            580                 585                 590

acc aca aca cca ggc ctc agt gag gca tct acc acc ttc tac agc agc   1824
Thr Thr Thr Pro Gly Leu Ser Glu Ala Ser Thr Thr Phe Tyr Ser Ser
        595                 600                 605

ccc aga tca cca acc aca aca ctc tca cct gcc agc atg aca agc cta   1872
Pro Arg Ser Pro Thr Thr Thr Leu Ser Pro Ala Ser Met Thr Ser Leu
    610                 615                 620

ggc gtc ggt gaa gaa tcc acc acc tcc cgt agc caa cca ggt tct act   1920
Gly Val Gly Glu Glu Ser Thr Thr Ser Arg Ser Gln Pro Gly Ser Thr
625                 630                 635                 640

cac tca aca gtg tca cct gcc agc acc acc acg cca ggc ctc agt gag   1968
His Ser Thr Val Ser Pro Ala Ser Thr Thr Thr Pro Gly Leu Ser Glu
                645                 650                 655

gaa tct acc acc gtc tac agc agc agc cca ggc tca act gaa acc aca   2016
Glu Ser Thr Thr Val Tyr Ser Ser Ser Pro Gly Ser Thr Glu Thr Thr
            660                 665                 670

gtg ttc cct cgc agc acc aca acc tca gtt cgt ggt gaa gag cct aca   2064
Val Phe Pro Arg Ser Thr Thr Thr Ser Val Arg Gly Glu Glu Pro Thr
        675                 680                 685

acc ttc cac agc cgg cca gcc tca act cac aca aca ctg ttc act gag   2112
Thr Phe His Ser Arg Pro Ala Ser Thr His Thr Thr Leu Phe Thr Glu
    690                 695                 700

gac agc acc acc tcg ggc ctc act gaa gaa tct aca gcc ttc ccc ggc   2160
Asp Ser Thr Thr Ser Gly Leu Thr Glu Glu Ser Thr Ala Phe Pro Gly
705                 710                 715                 720

agc cca gcc tcc acc caa aca ggg tta cct gcc aca ctc aca acc gca   2208
Ser Pro Ala Ser Thr Gln Thr Gly Leu Pro Ala Thr Leu Thr Thr Ala
                725                 730                 735

gac ctc ggt gag gaa tca act acc ttt ccc agc agc tca ggc tca act   2256
Asp Leu Gly Glu Glu Ser Thr Thr Phe Pro Ser Ser Ser Gly Ser Thr
            740                 745                 750

gga aca aca ctc tca cct gcc cgc tcc acc acc tct ggc ctc gtt gga   2304
Gly Thr Thr Leu Ser Pro Ala Arg Ser Thr Thr Ser Gly Leu Val Gly
        755                 760                 765

                                                                  2352
```

```
Glu Ser Thr Pro Ser Arg Leu Ser Pro Ser Ser Thr Glu Thr Thr Thr
    770                 775                 780

tta ccc ggc agt ccc aca aca cca agc ctc agt gag aaa tca acc acc   2400
Leu Pro Gly Ser Pro Thr Thr Pro Ser Leu Ser Glu Lys Ser Thr Thr
785                 790                 795                 800

ttc tac act agc ccc aga tca cca gat gca aca ctc tca cct gca acc   2448
Phe Tyr Thr Ser Pro Arg Ser Pro Asp Ala Thr Leu Ser Pro Ala Thr
                805                 810                 815

aca aca agc tca ggc gtc agc gaa gaa tcc agc aca tcc cac agt caa   2496
Thr Thr Ser Ser Gly Val Ser Glu Glu Ser Ser Thr Ser His Ser Gln
            820                 825                 830

cca ggc tca acg cac aca aca gcg ttc cct gac agc acc acc acc tca   2544
Pro Gly Ser Thr His Thr Thr Ala Phe Pro Asp Ser Thr Thr Thr Ser
        835                 840                 845

ggc ctc agt cag gaa cct aaa act tcc cac agc agc caa ggc tca aca   2592
Gly Leu Ser Gln Glu Pro Lys Thr Ser His Ser Ser Gln Gly Ser Thr
    850                 855                 860

gag gca aca ctg tcc cct ggc agt acc aca gcc tca tcc ctt ggt caa   2640
Glu Ala Thr Leu Ser Pro Gly Ser Thr Thr Ala Ser Ser Leu Gly Gln
865                 870                 875                 880

caa tct aca acc ttc cac agc agc cca ggc gac act gaa acc aca ctc   2688
Gln Ser Thr Thr Phe His Ser Ser Pro Gly Asp Thr Glu Thr Thr Leu
                885                 890                 895

tta cct gat gac acc ata acc tca ggc ctc gtg gag gca tct aca ccc   2736
Leu Pro Asp Asp Thr Ile Thr Ser Gly Leu Val Glu Ala Ser Thr Pro
            900                 905                 910

acc cac agc agc act ggc tcg cta cac aca aca ctg acc cct gcc agc   2784
Thr His Ser Ser Thr Gly Ser Leu His Thr Thr Leu Thr Pro Ala Ser
        915                 920                 925

tcc aca agc gct ggc ctt cag gaa gaa tct act act ttc cag agc tgg   2832
Ser Thr Ser Ala Gly Leu Gln Glu Glu Ser Thr Thr Phe Gln Ser Trp
    930                 935                 940

cca agc tca agt gac aca aca cct tca cct ccc ggc ccg g             2872
Pro Ser Ser Ser Asp Thr Thr Pro Ser Pro Pro Gly Pro
945                 950                 955
```

<210> 3
<211> 957
<212> PRT
<213> Homo sapiens

<400> 3

```
Arg Asn Arg Pro His Thr Thr Ala Phe Pro Gly Ser Thr Thr Met Pro
 1               5                  10                  15

Gly Val Ser Gln Glu Ser Thr Ala Ser His Ser Ser Pro Gly Ser Thr
            20                  25                  30

Asp Thr Thr Leu Ser Pro Gly Ser Thr Thr Ala Ser Ser Leu Gly Pro
        35                  40                  45

Glu Ser Thr Thr Phe His Ser Gly Pro Gly Ser Thr Glu Thr Thr Leu
    50                  55                  60

Leu Pro Asp Asn Thr Thr Ala Ser Gly Leu Leu Glu Ala Ser Thr Pro
65                  70                  75                  80

Val His Ser Ser Thr Gly Ser Pro His Thr Thr Leu Ser Pro Ala Gly
                85                  90                  95

Ser Thr Thr Arg Gln Gly Glu Ser Thr Thr Phe Gln Ser Trp Pro Asn
            100                 105                 110

Ser Lys Asp Thr Thr Pro Ala Pro Pro Thr Thr Thr Ser Ala Phe Val
        115                 120                 125

Glu Leu Ser Thr Thr Ser His Gly Ser Pro Ser Ser Thr Pro Thr Thr
    130                 135                 140

His Phe Ser Ala Ser Ser Thr Thr Leu Gly Arg Ser Glu Glu Ser Thr
145                 150                 155                 160

Thr Val His Ser Ser Pro Val Ala Thr Ala Thr Thr Pro Ser Pro Ala
                165                 170                 175

Arg Ser Thr Thr Ser Gly Leu Val Glu Glu Ser Thr Thr Tyr His Ser
            180                 185                 190

Ser Pro Gly Ser Thr Gln Thr Met His Phe Pro Glu Ser Asp Thr Thr
        195                 200                 205

Ser Gly Arg Gly Glu Glu Ser Thr Thr Ser His Ser Ser Thr Thr His
    210                 215                 220
```

Thr Ile Ser Ser Ala Pro Ser Thr Thr Ser Ala Leu Val Glu Glu Pro
225 230 235 240

Thr Ser Tyr His Ser Ser Pro Gly Ser Thr Ala Thr Thr His Phe Pro
245 250 255

Asp Ser Ser Thr Thr Ser Gly Arg Ser Glu Glu Ser Thr Ala Ser His
260 265 270

Ser Asn Gln Asp Ala Thr Gly Thr Ile Val Leu Pro Ala Arg Ser Thr
275 280 285

Thr Ser Val Leu Leu Gly Glu Ser Thr Thr Ser Pro Ile Ser Ser Gly
290 295 300

Ser Met Glu Thr Thr Ala Leu Pro Gly Ser Thr Thr Thr Pro Gly Leu
305 310 315 320

Ser Glu Lys Ser Thr Thr Phe His Ser Ser Pro Arg Ser Pro Ala Thr
325 330 335

Thr Leu Ser Pro Ala Ser Thr Thr Ser Ser Gly Val Ser Glu Glu Ser
340 345 350

Thr Thr Ser His Ser Arg Pro Gly Ser Thr His Thr Thr Ala Phe Pro
355 360 365

Asp Ser Thr Thr Thr Pro Gly Leu Ser Arg His Ser Thr Thr Ser His
370 375 380

Ser Ser Pro Gly Ser Thr Asp Thr Thr Leu Leu Pro Ala Ser Thr Thr
385 390 395 400

Thr Ser Gly Pro Ser Gln Glu Ser Thr Thr Ser His Ser Ser Pro Gly
405 410 415

Ser Thr Asp Thr Ala Leu Ser Pro Gly Ser Thr Thr Ala Leu Ser Phe
420 425 430

Gly Gln Glu Ser Thr Thr Phe His Ser Ser Pro Gly Ser Thr His Thr
435 440 445

Thr Leu Phe Pro Asp Ser Thr Thr Ser Ser Gly Ile Val Glu Ala Ser
450 455 460

Thr Arg Val His Ser Ser Thr Gly Ser Pro Arg Thr Thr Leu Ser Pro
465 470 475 480

Ala Ser Ser Thr Ser Pro Gly Leu Gln Gly Glu Ser Thr Ala Phe Gln
485 490 495

Thr His Pro Ala Ser Thr His Thr Thr Pro Ser Thr Pro Ser Thr Ala
500 505 510

Thr Ala Pro Val Glu Glu Ser Thr Thr Tyr His Arg Ser Pro Ser Ser
515 520 525

Thr Pro Thr Thr His Phe Pro Ala Ser Ser Thr Thr Ser Gly His Ser
530 535 540

Glu Lys Ser Thr Ile Phe His Ser Ser Pro Asp Ala Ser Gly Thr Thr
545 550 555 560

Pro Ser Ser Ala His Ser Thr Thr Ser Gly Arg Gly Glu Ser Thr Thr
565 570 575

Ser Arg Ile Ser Pro Gly Ser Thr Glu Ile Thr Thr Leu Pro Gly Ser
580 585 590

Thr Thr Thr Pro Gly Leu Ser Glu Ala Ser Thr Thr Phe Tyr Ser Ser
595 600 605

Pro Arg Ser Pro Thr Thr Thr Leu Ser Pro Ala Ser Met Thr Ser Leu
610 615 620

Gly Val Gly Glu Glu Ser Thr Thr Ser Arg Ser Gln Pro Gly Ser Thr
625 630 635 640

His Ser Thr Val Ser Pro Ala Ser Thr Thr Thr Pro Gly Leu Ser Glu
645 650 655

Glu Ser Thr Thr Val Tyr Ser Ser Ser Pro Gly Ser Thr Glu Thr Thr
660 665 670

Val Phe Pro Arg Ser Thr Thr Thr Ser Val Arg Gly Glu Glu Pro Thr
675 680 685

Thr Phe His Ser Arg Pro Ala Ser Thr His Thr Thr Leu Phe Thr Glu
690 695 700

Asp Ser Thr Thr Ser Gly Leu Thr Glu Glu Ser Thr Ala Phe Pro Gly
705 710 715 720

Ser Pro Ala Ser Thr Gln Thr Gly Leu Pro Ala Thr Leu Thr Thr Ala
725 730 735

Asp Leu Gly Glu Glu Ser Thr Thr Phe Pro Ser Ser Ser Gly Ser Thr
740 745 750

Gly Thr Thr Leu Ser Pro Ala Arg Ser Thr Thr Ser Gly Leu Val Gly
755 760 765

Glu Ser Thr Pro Ser Arg Leu Ser Pro Ser Ser Thr Glu Thr Thr Thr
770 775 780

Leu Pro Gly Ser Pro Thr Thr Pro Ser Leu Ser Glu Lys Ser Thr Thr
785 790 795 800

Phe Tyr Thr Ser Pro Arg Ser Pro Asp Ala Thr Leu Ser Pro Ala Thr
805 810 815

Thr Thr Ser Ser Gly Val Ser Glu Glu Ser Ser Thr Ser His Ser Gln
820 825 830

Pro Gly Ser Thr His Thr Thr Ala Phe Pro Asp Ser Thr Thr Thr Ser
835 840 845

Gly Leu Ser Gln Glu Pro Lys Thr Ser His Ser Ser Gln Gly Ser Thr
850 855 860

Glu Ala Thr Leu Ser Pro Gly Ser Thr Thr Ala Ser Ser Leu Gly Gln
865 870 875 880

Gln Ser Thr Thr Phe His Ser Ser Pro Gly Asp Thr Glu Thr Thr Leu
885 890 895

Leu Pro Asp Asp Thr Ile Thr Ser Gly Leu Val Glu Ala Ser Thr Pro
900 905 910

Thr His Ser Ser Thr Gly Ser Leu His Thr Thr Leu Thr Pro Ala Ser
915 920 925

Ser Thr Ser Ala Gly Leu Gln Glu Glu Ser Thr Thr Phe Gln Ser Trp
930 935 940

Pro Ser Ser Ser Asp Thr Thr Pro Ser Pro Pro Gly Pro
945 950 955

<210> 4
<211> 28
<212> PRT
<213> Homo sapiens

<220>
<221> REPEAT
<222> (1)..(28)
<223> MUC12 consensus tandem repeat sequence

<400> 4
Ser Gly Leu Ser Gln Glu Ser Thr Thr Phe His Ser Ser Pro Gly Ser
1 5 10 15

Thr Glu Thr Thr Leu Ser Pro Ala Ser Thr Thr Thr
20 25

<210> 5
<211> 2095
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (3)..(1757)

<400> 5
aa aca ctc tca cct gcc agc atg aga agc tcc agc atc agt gga gaa 47
Thr Leu Ser Pro Ala Ser Met Arg Ser Ser Ser Ile Ser Gly Glu
1 5 10 15

ccc acc agc ttg tat agc caa gca gag tca aca cac aca aca gcg ttc 95
Pro Thr Ser Leu Tyr Ser Gln Ala Glu Ser Thr His Thr Thr Ala Phe
20 25 30

cct gcc agc acc acc acc tca ggc ctc agt cag gaa tca aca act ttc 143
Pro Ala Ser Thr Thr Thr Ser Gly Leu Ser Gln Glu Ser Thr Thr Phe
35 40 45

cac agt aag cca ggc tca act gag aca aca ctg tcc cct ggc agc atc 191
His Ser Lys Pro Gly Ser Thr Glu Thr Thr Leu Ser Pro Gly Ser Ile
50 55 60

aca act tca tct ttt gct caa gaa ttt acc acc cct cat agc caa cca 239
Thr Thr Ser Ser Phe Ala Gln Glu Phe Thr Thr Pro His Ser Gln Pro
65 70 75

ggc tca gct ctg tca aca gtg tca cct gcc agc acc aca gtg cca ggc 287
Gly Ser Ala Leu Ser Thr Val Ser Pro Ala Ser Thr Thr Val Pro Gly
80 85 90 95

335

```
Leu Ser Glu Glu Ser Thr Thr Phe Tyr Ser Ser Pro Gly Ser Thr Glu
                100                 105                 110

acc aca gcg ttt tct cac agc aac aca atg tcc att cat agt caa caa   383
Thr Thr Ala Phe Ser His Ser Asn Thr Met Ser Ile His Ser Gln Gln
            115                 120                 125

tct aca ccc ttc cct gac agc cca ggc ttc act cac aca gtg tta cct   431
Ser Thr Pro Phe Pro Asp Ser Pro Gly Phe Thr His Thr Val Leu Pro
        130                 135                 140

gcc acc ctc aca acc aca gac att ggt cag gaa tca aca gcc ttc cac   479
Ala Thr Leu Thr Thr Thr Asp Ile Gly Gln Glu Ser Thr Ala Phe His
    145                 150                 155

agc agc tca gac gca act gga aca aca ccc tta cct gcc cgc tcc aca   527
Ser Ser Ser Asp Ala Thr Gly Thr Thr Pro Leu Pro Ala Arg Ser Thr
160                 165                 170                 175

gcc tca gac ctt gtt gga gaa cct aca act ttc tac atc agc cca tcc   575
Ala Ser Asp Leu Val Gly Glu Pro Thr Thr Phe Tyr Ile Ser Pro Ser
                180                 185                 190

cct act tac aca aca ctc ttt cct gcg agt tcc agc aca tca ggc ctc   623
Pro Thr Tyr Thr Thr Leu Phe Pro Ala Ser Ser Ser Thr Ser Gly Leu
            195                 200                 205

act gag gaa tct acc acc ttc cac acc agt cca agc ttc act tct aca   671
Thr Glu Glu Ser Thr Thr Phe His Thr Ser Pro Ser Phe Thr Ser Thr
        210                 215                 220

att gtg tct act gaa agc ctg gaa acc tta gca cca ggg ttg tgc cag   719
Ile Val Ser Thr Glu Ser Leu Glu Thr Leu Ala Pro Gly Leu Cys Gln
    225                 230                 235

gaa gga caa att tgg aat gga aaa caa tgc gtc tgt ccc caa ggc tac   767
Glu Gly Gln Ile Trp Asn Gly Lys Gln Cys Val Cys Pro Gln Gly Tyr
240                 245                 250                 255

gtt ggt tac cag tgc ttg tcc cct ctg gaa tcc ttc cct gta gaa acc   815
Val Gly Tyr Gln Cys Leu Ser Pro Leu Glu Ser Phe Pro Val Glu Thr
                260                 265                 270

ccg gaa aaa ctc aac gcc act tta ggt atg aca gtg aaa gtg act tac   863
Pro Glu Lys Leu Asn Ala Thr Leu Gly Met Thr Val Lys Val Thr Tyr
            275                 280                 285
```

```
Arg Asn Phe Thr Glu Lys Met Asn Asp Ala Ser Ser Gln Glu Tyr Gln
        290                 295                 300

aac ttc agt acc ctc ttc aag aat cgg atg gat gtc gtt ttg aag ggc   959
Asn Phe Ser Thr Leu Phe Lys Asn Arg Met Asp Val Val Leu Lys Gly
    305                 310                 315

gac aat ctt cct cag tat aga ggg gtg aac att cgg aga ttg ctc aac   1007
Asp Asn Leu Pro Gln Tyr Arg Gly Val Asn Ile Arg Arg Leu Leu Asn
320                 325                 330                 335

ggt agc atc gtg gtc aag aac gat gtc atc ctg gag gca gac tac act   1055
Gly Ser Ile Val Val Lys Asn Asp Val Ile Leu Glu Ala Asp Tyr Thr
                340                 345                 350

tta gag tat gag gaa ctg ttt gaa aac ctg gca gag att gta aag gcc   1103
Leu Glu Tyr Glu Glu Leu Phe Glu Asn Leu Ala Glu Ile Val Lys Ala
            355                 360                 365

aag att atg aat gaa act aga aca act ctt ctt gat cct gat tcc tgc   1151
Lys Ile Met Asn Glu Thr Arg Thr Thr Leu Leu Asp Pro Asp Ser Cys
        370                 375                 380

aga aag gcc ata ctg tgc tat agt gaa gag gac act ttc gtg gat tca   1199
Arg Lys Ala Ile Leu Cys Tyr Ser Glu Glu Asp Thr Phe Val Asp Ser
    385                 390                 395

tcg gtg act ccg ggc ttt gac ttc cag gag caa tgc acc cag aag gct   1247
Ser Val Thr Pro Gly Phe Asp Phe Gln Glu Gln Cys Thr Gln Lys Ala
400                 405                 410                 415

gcc gaa gga tat acc cag ttc tac tat gtg gat gtc ttg gat ggg aag   1295
Ala Glu Gly Tyr Thr Gln Phe Tyr Tyr Val Asp Val Leu Asp Gly Lys
                420                 425                 430

ctg gcc tgt gtg aac aag tgc acc aaa gga acg aag tcg caa atg aac   1343
Leu Ala Cys Val Asn Lys Cys Thr Lys Gly Thr Lys Ser Gln Met Asn
            435                 440                 445

tgt aac ctg ggc aca tgt cag ctg caa cgc agt ggc ccc cgc tgc ctg   1391
Cys Asn Leu Gly Thr Cys Gln Leu Gln Arg Ser Gly Pro Arg Cys Leu
        450                 455                 460

tgc cca aat acg aac aca cac tgg tac tgg gga gag acc tgt gaa ttc   1439
Cys Pro Asn Thr Asn Thr His Trp Tyr Trp Gly Glu Thr Cys Glu Phe
    465                 470                 475

                                                                  1487
```

```
Asn Ile Ala Lys Ser Leu Val Tyr Gly Ile Val Gly Ala Val Met Ala
480                 485                 490                 495

gtg ctg ctg ctc gca ttg atc atc cta atc atc tta ttc agc cta tcc   1535
Val Leu Leu Leu Ala Leu Ile Ile Leu Ile Ile Leu Phe Ser Leu Ser
                500                 505                 510

cag aga aaa cgg cac agg gaa cag tat gat gtg cct caa gag tgg cga   1583
Gln Arg Lys Arg His Arg Glu Gln Tyr Asp Val Pro Gln Glu Trp Arg
            515                 520                 525

aag gaa ggc acc cct ggc atc ttc cag aag acg gcc atc tgg gaa gac   1631
Lys Glu Gly Thr Pro Gly Ile Phe Gln Lys Thr Ala Ile Trp Glu Asp
        530                 535                 540

cag aat ctg agg gag agc aga ttc ggc ctt gag aac gcc tac aac aac   1679
Gln Asn Leu Arg Glu Ser Arg Phe Gly Leu Glu Asn Ala Tyr Asn Asn
    545                 550                 555

ttc cgg ccc acc ctg gag act gtt gac tct ggc aca gag ctc cac atc   1727
Phe Arg Pro Thr Leu Glu Thr Val Asp Ser Gly Thr Glu Leu His Ile
560                 565                 570                 575

cag agg ccg gag atg gta gca tcc act gtg tgagccaacg ggggcctccc     1777
Gln Arg Pro Glu Met Val Ala Ser Thr Val
                580                 585

accctcatct agctctgttc aggagagctg caaacacaga gcccaccaca agcctccggg 1837

gcgggtcaag aggagaccga agtcaggccc tgaagccggt cctgctctga gctgacagac 1897

ttggccagtc ccctgcctgt gctcctgctg gggaaggctg ggggctgtaa gcctctccat 1957

ccgggagctt ccagactccc agaagcctcg gcacccctgt ctcctcctgg gtggctcccc 2017

actctggaat ttccctacca ataaaagcaa atctgaaagc tcaaaaaaaa aaaaaaaaaa 2077

aaaaaaaaaa aaaaaaaa                                               2095
```

<210> 6
<211> 585
<212> PRT
<213> Homo sapiens

<400> 6
Thr Leu Ser Pro Ala Ser Met Arg Ser Ser Ser Ile Ser Gly Glu Pro

Thr Ser Leu Tyr Ser Gln Ala Glu Ser Thr His Thr Thr Ala Phe Pro
20 25 30

Ala Ser Thr Thr Thr Ser Gly Leu Ser Gln Glu Ser Thr Thr Phe His
35 40 45

Ser Lys Pro Gly Ser Thr Glu Thr Thr Leu Ser Pro Gly Ser Ile Thr
50 55 60

Thr Ser Ser Phe Ala Gln Glu Phe Thr Thr Pro His Ser Gln Pro Gly
65 70 75 80

Ser Ala Leu Ser Thr Val Ser Pro Ala Ser Thr Thr Val Pro Gly Leu
85 90 95

Ser Glu Glu Ser Thr Thr Phe Tyr Ser Ser Pro Gly Ser Thr Glu Thr
100 105 110

Thr Ala Phe Ser His Ser Asn Thr Met Ser Ile His Ser Gln Gln Ser
115 120 125

Thr Pro Phe Pro Asp Ser Pro Gly Phe Thr His Thr Val Leu Pro Ala
130 135 140

Thr Leu Thr Thr Thr Asp Ile Gly Gln Glu Ser Thr Ala Phe His Ser
145 150 155 160

Ser Ser Asp Ala Thr Gly Thr Thr Pro Leu Pro Ala Arg Ser Thr Ala
165 170 175

Ser Asp Leu Val Gly Glu Pro Thr Thr Phe Tyr Ile Ser Pro Ser Pro
180 185 190

Thr Tyr Thr Thr Leu Phe Pro Ala Ser Ser Ser Thr Ser Gly Leu Thr
195 200 205

Glu Glu Ser Thr Thr Phe His Thr Ser Pro Ser Phe Thr Ser Thr Ile
210 215 220

Val Ser Thr Glu Ser Leu Glu Thr Leu Ala Pro Gly Leu Cys Gln Glu
225 230 235 240

Gly Gln Ile Trp Asn Gly Lys Gln Cys Val Cys Pro Gln Gly Tyr Val
245 250 255

Gly Tyr Gln Cys Leu Ser Pro Leu Glu Ser Phe Pro Val Glu Thr Pro
270

Glu Lys Leu Asn Ala Thr Leu Gly Met Thr Val Lys Val Thr Tyr Arg
275 280 285

Asn Phe Thr Glu Lys Met Asn Asp Ala Ser Ser Gln Glu Tyr Gln Asn
290 295 300

Phe Ser Thr Leu Phe Lys Asn Arg Met Asp Val Val Leu Lys Gly Asp
305 310 315 320

Asn Leu Pro Gln Tyr Arg Gly Val Asn Ile Arg Arg Leu Leu Asn Gly
325 330 335

Ser Ile Val Val Lys Asn Asp Val Ile Leu Glu Ala Asp Tyr Thr Leu
340 345 350

Glu Tyr Glu Glu Leu Phe Glu Asn Leu Ala Glu Ile Val Lys Ala Lys
355 360 365

Ile Met Asn Glu Thr Arg Thr Thr Leu Leu Asp Pro Asp Ser Cys Arg
370 375 380

Lys Ala Ile Leu Cys Tyr Ser Glu Glu Asp Thr Phe Val Asp Ser Ser
385 390 395 400

Val Thr Pro Gly Phe Asp Phe Gln Glu Gln Cys Thr Gln Lys Ala Ala
405 410 415

Glu Gly Tyr Thr Gln Phe Tyr Tyr Val Asp Val Leu Asp Gly Lys Leu
420 425 430

Ala Cys Val Asn Lys Cys Thr Lys Gly Thr Lys Ser Gln Met Asn Cys
435 440 445

Asn Leu Gly Thr Cys Gln Leu Gln Arg Ser Gly Pro Arg Cys Leu Cys
450 455 460

Pro Asn Thr Asn Thr His Trp Tyr Trp Gly Glu Thr Cys Glu Phe Asn
465 470 475 480

Ile Ala Lys Ser Leu Val Tyr Gly Ile Val Gly Ala Val Met Ala Val
485 490 495

Leu Leu Leu Ala Leu Ile Ile Leu Ile Ile Leu Phe Ser Leu Ser Gln
500 505 510

Arg Lys Arg His Arg Glu Gln Tyr Asp Val Pro Gln Glu Trp Arg Lys

Glu Gly Thr Pro Gly Ile Phe Gln Lys Thr Ala Ile Trp Glu Asp Gln
530 535 540

Asn Leu Arg Glu Ser Arg Phe Gly Leu Glu Asn Ala Tyr Asn Asn Phe
545 550 555 560

Arg Pro Thr Leu Glu Thr Val Asp Ser Gly Thr Glu Leu His Ile Gln
565 570 575

Arg Pro Glu Met Val Ala Ser Thr Val
580 585

<210> 7
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Random 10mer PCR primer

<400> 7
acttcgccac 10

<210> 8
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:MUC 12 PCR forward primer

<400> 8
tgaagggcga caatcttcct c 21

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:MUC12 reverse

<400> 9
tacacgaggc tcttggcgat gttg 24

<210> 10
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Mucl1 forward
PCR primer

<400> 10
caggcgtcag tcaggaatct acag 24

<210> 11
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Mucl1 reverse
PCR primer

<400> 11
gaggctgtgg tgttgtcagg taag 24

<210> 12
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Beta2-microglobulin forward PCR primer

<400> 12
tgaattgcta tgtgtctggg t 21

<210> 13
<211> 21
<212> DNA

<220>
<223> Description of Artificial Sequence: Beta
2-microglobulin reverse PCR primer

<400> 13
cctccatgat gctgcttaca t 21

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: MUC12 forward
primer for verification of contiguous sequence

<400> 14
agccaaccag gctcagctct 20

<210> 15
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: MUC12 reverse
primer for verification of contiguous sequence

<400> 15
gctcacacag tggatgctac c 21

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Clone Ii5
reverse PCR primer

<400> 16
gggaacactg tggtttcagt tgag 24

<210> 17
<211> 17
<212> PRT
<213> Homo sapiens

<220>
<221> PEPTIDE
<222> (1)..(17)
<223> MUC11 immunizing peptide

<400> 17
Cys Phe His Ser Arg Pro Ala Ser Thr His Thr Thr Leu Phe Thr Glu
1 5 10 15

Asp

<210> 18
<211> 17
<212> PRT
<213> Homo sapiens

<220>
<221> PEPTIDE
<222> (1)..(17)
<223> MUC12 immunizing peptide

<400> 18
Thr Tyr Arg Asn Phe Thr Glu Lys Met Asn Asp Ala Ser Ser Gln Glu
1 5 10 15

Cys

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 99/00579

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl[6]: C12N 15/12 C07K 14/435

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
ORBIT: C12N/IC and MUCIN#
MEDLINE: Mucins, Genetics/CT(L)Genetics AND (Chromosome Mapping/CT OR Chromosomes/CT)
CA: Mucins/CT AND (Gene Structure/IT OR Chromosome/IT OR Chromosomal/IT)
SWISSPROT, EMBL, PIR: SEQ IDs 1 - 6

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Fox, M. F. et al *Ann. Hum. Genet.* (1992) 56, pages 281 - 287 "Regional localizational of the intestinal mucin gene MUC3 to chromosome 7q22" Whole Document	1
X	Gum, J. R. et al *Biochem Biophys Res Commun* (1990), 171(1) pages 407 - 415 "Molecular Cloning of cDNAs Derived from a Novel Human Intestinal Mucin Gene" Whole Document, page 412 last paragraph	1

☐ Further documents are listed in the continuation of Box C

☐ See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
12 August 1999

Date of mailing of the international search report
13 AUG 1999

Name and mailing address of the ISA/AU
AUSTRALIAN PATENT OFFICE
PO BOX 200
WODEN ACT 2606
AUSTRALIA

Authorized officer

ALISTAIR BESTOW
Telephone No : (02) 6283 2450